Handwritten Malayalam Character Recognition using Curvelet Transform and ANN

Manju Manuel
Associate Professor,
Dept. of ECE
R.I.T KOTTAYAM

Saidas S. R.
M.Tech. Scholar,
Dept. of ECE
GCE KANNUR

ABSTRACT
Malayalam, the official language of Kerala, a southern state of India has been accorded the honour of language of eminence. Hence the researches in recognition and related works in Malayalam language is gaining more prominence in the current scenario. This paper proposes the use of Curvelet transform and neural network for the recognition of handwritten Malayalam character. Curvelet transform is to be used in the feature extraction stage and neural network for classification. Curvelet transform provides a compact representation for curved singularities and is well suited for malayalam language. Two different back propagation algorithms had been employed and the performance is compared on varying architecture. The promising feature of the work is successful classification of 53 characters which is an improvement over the existing works. Application of character recognition include sorting of bank cheques and postal letters, reading aid for blind, data compression etc. Besides, an automated tool with graphical user interface in MATLAB has been developed for Malayalam character recognition.

General Terms
Pattern Recognition, Artificial Neural Network (ANN), Curvelet Transform, Optical character recognition (OCR),

Keywords
Malayalam, Character Recognition, Artificial Neural Network (ANN), Curvelet Transform, Handwritten.

1. INTRODUCTION
The overwhelming volume of paper-based information on companies and offices challenges their ability to manage documents and records. Since computers can work much faster andmore efficiently than human. It is used to perform many of the tasks required for efficient document and content management. But computer knows only alphanumeric characters as ASCIIcode. So computer cannot distinguish character or a word from a scanned image. In order to use the computer for document management and automatic sorting of mails etc., it isrequired to retrieve alphanumeric information from a scanned image. Optical character recognition system (OCR) allows us to convert a document into electronic text, which can edit andsearch etc. At present, more sophisticated optical readers are available for English, Roman, Chinese, Japanese and Arabic text. In the United States, about 60% of the hand-printed is sorted automatically. At the same time the OCR in Malayalam is at the beginning stage.

Malayalam is a Dravidian language with about 35 million speakers. It is spoken mainly in the south western India, particularly in Kerala. Until the 16^{th} century Malayalam was written in the vattezhuthu script. Modern Malayalam script is derived from the Grantha script, a descendant of the ancient Brahamiscript.Malayalam is written from left to right. Thereare fifty-three letters, called akaras [1]. As a result of the difficulties of printing Malayalam, a simplified or reformed version of the script was introduced during the 1970s and 1980s. The main change involved writing consonants and diacritics separately rather than as complex characters. These changes has not been consistently so the modern script is often a mixture of traditional and simplified letters [2]. The character set consists of 13 vowels, 2 left vowel signs, 7 right vowel signs, some appear on both sides of the Conj/consonant, 30 commonly used conjuncts.

Character recognition or optical character recognition (OCR) is the conversion of scanned or photographed images of typewritten or printed text into computer readable text. Optical character recognition has become one of the most successful applications of technology in the field of pattern recognition and artificial intelligence. Character recognition has a wide range of applications in banking and post office etc.

1.1 Motivation
Now days Information Technology based applications are growing rapidly and with the increase in these applications there is need of character recognition. Applications of character recognition are sorting of bank cheques and postal letters, reading aid for blind, Data compression etc.

In the field of Malayalam optical character Recognition very few works have been reported so far. One reason for this Malayalam language is rich in patterns while the combinations of such patterns make the problem even more complex

1.2 Optical Character Recognition
Optical Character Recognition (OCR) is the one of the most successful application in the field of pattern recognition and artificial intelligence. It deals with recognizing optically processed character. Optical character can be performed either in off-line or on-line. In off-line method recognition is done after writing or printing has been completed while in on-line recognition is done at the same time as character is drown or print. For these systems to effectively recognize hand-printed or machine- printed forms, individual characters must be well separated.

Fig. 1 shows different areas of character recognition [3]. The performance of an OCR is mainly depend on quality of input document. With regard to recognition accuracy, constrained or printed character, have higher performance than for unconstrained character.

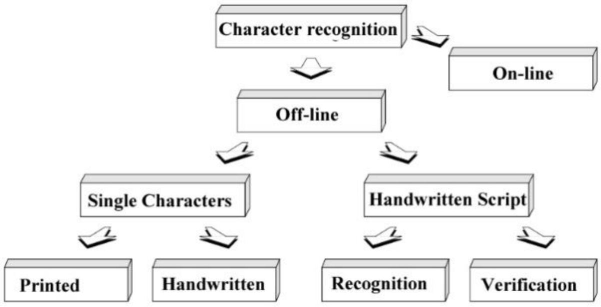


Fig. 1. The different areas of character recognition

2. MALAYALAM CHARACTER RECOGNITION -A REVIEW

Some of the existing techniques used in OCR for Malayalam, the official language of Kerala a state in India. NAYANA™ OCR developed by CDAC is based on theFeature Extraction method of character recognition. It uses anOtsus algorithm based thresholding approach for convertinggrey scale image to binary. Skew detection is done usingthe Projection profile. Another work for recognition of unconstrained isolated handwritten character recognition by G.Raju. In this paper samples are selected at 256 × 256 without any preprocessing such as denoising and thinning. Thenimage is converted to inverted binary and wavelet transformis applied using the wavelet DB4. For each sub band zerocrossing is used as feature for classification [4]. A NovelBilingual OCR for Printed Malayalam-English Text based onGabor Features and Dominant Singular Values is proposedby Bindu Philip R. D. Sudhaker Samuel. Which uses Gaborfeatures and twostage classification approach. They reporteda recoginition rate of 96.5% [5]. Another method by RenjuJohn and G. Raju and D. S. Guru it uses image of size 32× 64 and wavelet transform is used to find feature vector. 1-D Wavelet transform of horizontal and vertical projection isused as feature vector. A Multi-Level Perceptron network isused for classification [6]. Another work is done by M. AbdulRahiman and M. S. Rajasree uses horizontal and verticalscanning for separating lines in a document and for words ina line. For separating combinational letters like കാ labelingimage based method is used. Labeling is done by the four andeight connected pixels of an image for gettingക and ാ.Wavelet multi-resolution analysis for the purpose extractingfeatures, Daubechies (db4) wavelet is used. Artificial NeuralNetwork is used to accomplish the recognition tasks [7].

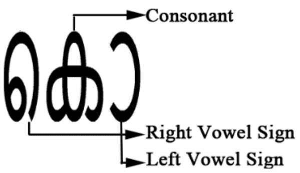


Fig. 2. A Typical Malayalam Character

Bindu Philip and R. D. Sudhaker Samuel works on An Efficient OCR for Printed Malayalam Text they used SVM classifiers and used a novel algorithm for segmentation. They considered thatprinted malayalam character have three segments as shown inFig. 2.The first segment could have either of one left vowel sign or two left vowel sign, the second segment it is core itcan be either vowel or consonant. The third segment is rightvowel sign. If we consider X as consonant and Y as vowel,0 as left vowel sign and 1 as right vowel sign we can haveX, Y, 0X, 00X, 0X1, X1 form of valid character sequences.For recognition five features are used and is classified usingSVM classifier [8]. Another work by Abdul Rahiman, M. S.Rajasree uses number of horizontal and vertical line as featurefor recognition of characters [9]. Ashitta T. Jia, YahkoobAyappally, and Syama K. uses N - gram segmentation approach andused along with geometric feature extraction. They uses thesefeature for SVM based classification [10]. Jomy John, PramodK. V, and KannanBalakrishnan proposed a method for offlinehandwritten Malayalam character recognition it uses chaincode and image centroid as a feature. Uses a median filterfor noise removal. Ostu's global thresholding for binarization[11].

3. THE OCR SYSTEM

Optical Character Recognition (OCR) is the conversion ofscanned or photoed images of typewritten or printed textinto machine-encoded/computer-readable text. A typical OCR system consists of several components. In Fig.3 a common setup is shown.

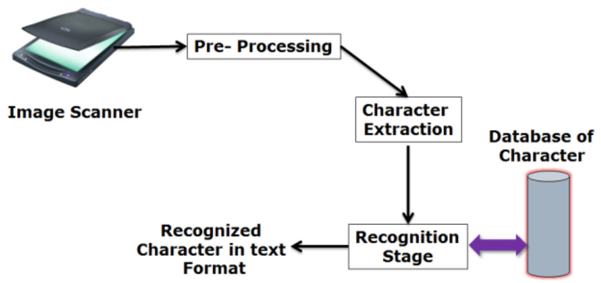


Fig. 3: Block diagram of typical character recognition system

Image scanning, Preprocessing, Character extraction,Recognition are the major steps involved in character recognition.

The first step in the process is to digitize the document.Optical scanners are used to digitize which generally consistof a transport mechanism plus a sensing device that convertslight intensity into gray-levels. Image scanning is done to getthe digital form of character on a page or paper, usually thescanning is done at a minimum of 300 dpi resolution. Theresolution at which the scanning is done plays a crucial rolein accuracy of character recognition.

The image resulting from the scanning process may containa certain amount of noise. Depending on the resolution onthe scanner and the success of the applied technique forthresholding, the characters may be smeared or broken. Someof these defects will later cause poor recognition rates. Thesenoises can be eliminated by using a preprocessor and wecan smooth the digitized characters. Preprocessing is doneto enhance the scanned images, by using binarization, noiseremoval, skew correction etc.

At Character recognition phase, first the line is extractedand word is extracted from the scanned document. Usually thehorizontal projection method and vertical projection methodsare used for extracting line and word in a document. In orderto detect each character, usually connected component methodis used. These extracted Characters used for feature extraction and further steps

Recognition phase involve two step:

1) Feature extraction and feature selection
2) Classification

Features chosen has to satisfy the condition 1) small intraclass variance 2) large inter-class variance. This means thatfeatures extracted from samples of the same class should besimilar, while that of different classes should be dissimilar.

These features are used for recognition stage. Recognition isachieved using artificial neural network. Before neural networkis used for recognition it has to be trained properly withcharacter database. After neural network is trained featuresfrom the character to be recognised are given to the inputof the neural network. From the previous training experiencethe neural network will classify the character to one of 53Character.

4. PROPOSED MALAYALAM CHARACTER RECOGNITION

The first step is the creation of a data base for Malayalam characters. Since no standard data base is available a set ofsamples are collected from persons having different age andsex group. A data base is created such a way that 53 charactersto be recognised are written over a white A4 size paper. 53characters consist of 8 vowels and 36 consonants. Vowels arenumbered from 1 to 8, consonant are numbered from 9 to 44and 45 to 53 are conjuncts in Fig. 4.

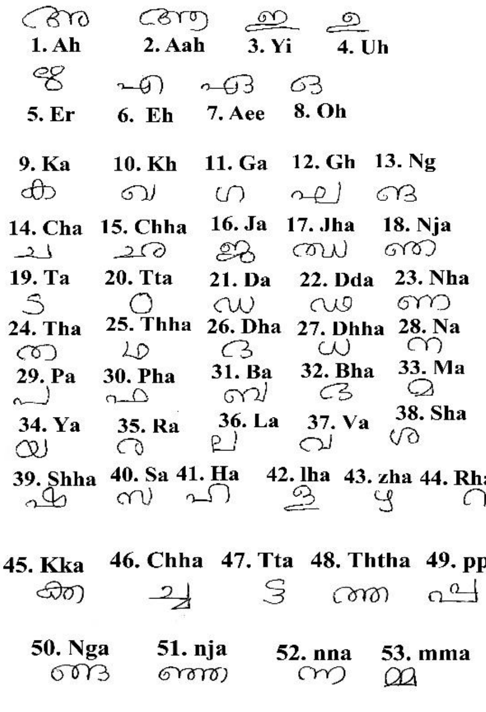


Fig. 4. Scanned Image of 53 character

These samples are scanned using Canon MF4820D. Scanning is done at a resolution of 300 - 400 dpi. Samples are madewith a white background. Then each character is segmentedand stored in different folders with some unique id's torepresent each character as shown Fig. 5. The folder nameMalayalam script database contains 53 folder corresponds theeach character. Each character folder contains 40 samples ofcharacter.

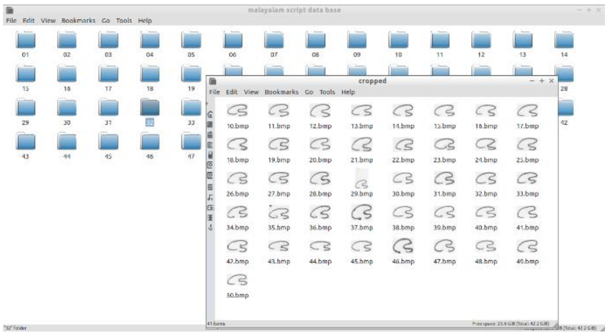


Fig. 5. Database of 53 character used

4.1 Preprocessing

For a scanned image it is observed that

1) The handwritten characters were concentrated in onlysome part of the image, the remaining being white space
2) The size of the image was too large to give as input toany of the classifier
3) Some images had noise in the example Fig. 6

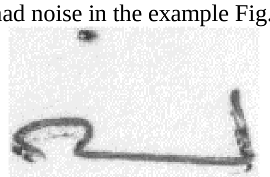


Fig. 6. Noise: A dot present in the sample of 'pa'

4) Characters in different images were of different size and concentrated in different parts of the image

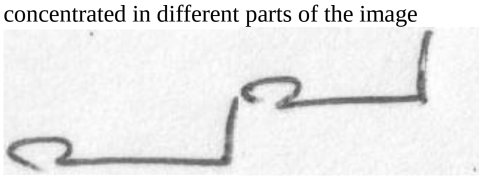


Fig. 7. First sample is bigger and concentrated in bottom part while second sample is smaller and concentrated in the top left

The above non uniformities in the samples would make theclassifiers perform badly. Pre-processing was used to normalize the samples so as to solve the above mentioned problem.Scanned image usually suffers from noise, usually Gaussiannoise and salt and pepper noise. Preprocessing is done tosuppress the effect of noise and to enhance quality of scannedimages.

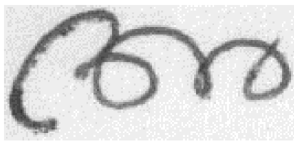


Fig. 8. Scanned image of malayalam letter 'ah'

The scanned image of Malayalam character 'Ah' is shownin Fig. 8 which is the intensity (Gray Scale) image and isfurther passes through following preprocessing stages.

- Noise Removal
- binarization
- Thinning
- Normalising the size of image

The noise present in the scanned images are Gaussian noiseand salt and pepper noise. Gaussian noise can be suppressedby median filtering.

The next step of preprocessing is binarization. The filteredscanned image is a Gray scale which has 256 intensity levels,which is converted to binary image having 2 level intensityie., white and black image using Ostu's global thresholdmethod. A fixed threshold is used, where gray-levels belowthis threshold is said to be black and levels above are said to bewhite. Fig. 9 shows the binary image after Global thresholdingmethod.

Fig. 9. binarized image of malayalam letter 'ah'

The binary image after thersholding has pixel value suchthat white pixel have logic '1' and black pixel has '0'. Forfurther steps the image is then inverted to make pixel valuecorresponds to the character to '1'.

Form the inverted image connected pixels less than somethreshold (say 30 pixels) is removed which corresponds tonoise as shown Fig. 10.

Fig. 10. invetedbinarized image of malayalam letter 'ah'

Further a bounding box is created which touches the character in four sides. It removes unnecessary black spaces asshown in Fig. 11

Fig. 11. Croped image

Then the image is resized to 32 × 64. Then perform thethinning operation is performed to make the each characterpixel to '1' - pixel width. Fig. 12 shows a thinned image ofthe Malayalam character 'ah'.

Fig. 12. Character after performing thinning operation

So, after preprocessing a thinned image of character resizedto 32 × 64 is obtained. Which is suite for further processing.

4.2 Feature Extraction

Feature extraction is the most important step in Characterrecognition. Feature extraction is a special form imensionalityreduction. Feature extraction reduces amount of resourcesrequired to describe a character. In this case the final outputis preprocessing of size 32 × 64 (=2048 bits) by finding out afeature we can represent the same character by lesser number of bits.

Feature extraction involve two steps

1) Obtaining projection profile
2) Computing transformation

In proposing method projection profile of 32 × 64 sizecharacter image taken. Basically two types of projectionsare used: horizontal and vertical. Projection profile is a datastructure used to store the number of non-background pixelwhen the image is projected over the normal X-Y axis. Eachcell of the projection vector is associated with the numberof pixels above a predefined threshold (usually backgroundcolor). A row vector of size 32 as horizontal projection profileand a row vector of size 64 as the vertical projection profileis obtained is shown in Fig. 13 .

$$Vertical\ Projection(c)=\sum_{r=1}^{32} A_{cr} \quad (1)$$

$$Horizontal\ Projection(c)=\sum_{c=1}^{64} A_{rc} \quad (2)$$

Where A is a black and white image with character havingwhite pixel. Value of r ranges from 1 to 32 and values of cranges from 1 to 64.

Fig. 13. Projection profile of typical character 'ah'

After projection profile of a character is obtained. Curvelet Transformations are used to computed features.

5. FEATURE EXTRACTION USING CURVELET TRANSFORM

The main contribution of the work is the use of curvelettransform for feature extraction. It had been chosen sinceMalayalam language possess many curved singularities andcurvelet provides an efficient representation. The Curvelettransform is a higher dimensional generalization of theWavelet transform. By using Curvelet Transform Curved singularities can be well approximated with very few coefficients [12]. Curvelet transform uses translations, dilations, and rotations of basic curvelet ψ. Curvelet transform has a highlyredundant dictionary which can provide sparse representationof signals that have edges along the regular curve. Conceptually, the curvelet transform is a multiscale pyramid withmany directions and positions at each length scale, and needle shaped elements at fine scales [13].

For the implementation of curvelet transform, first 2D Fast Fourier Transform (FFT) of the image data is taken. Then the 2D Fourier frequency plane is divided into wedges (like theshaded region in Fig. 14). The parabolic shape of wedges is the result of partitioning the Fourier plane into radial (concentric circles) and angular divisions.

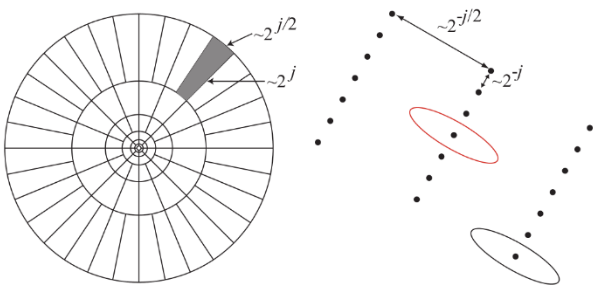


Fig. 14. Curvelets in Fourier frequency (left) and spatial domain (right)

The concentric circles areresponsible for the decomposition of an image into multiplescales (used for band passing the image at different scales)and the angular divisions partition the band passed image intodifferent angles or orientations. Thus if we want to deal witha particular wedge well need to define its scale j and angle l. Now let's have a look at the spatial domain (Fig. 14 right).Each of the wedges here corresponds to a particular curvelet(shown as ellipses) at a given scale and angle. This indicatesthat the inverse FFT of a particular wedge if taken, willdetermine the curvelet coefficients for that scale and angleCurvelet obey scalling low ie, length=width^2

There are two different ways of digital implementations ofFast Digital Curvelet Transform (FDCT).

1) Curvelets via USFFT (Unequally Spaced Fast FourierTransform)
2) Curvelets via Wrapping

In this work feature extraction is done using Curvelets viaUSFFT are used. Digital Curvelet transform is obtained by

$$C^D(j,k,l) := \sum_{0\le t_1,t_2} f[t_1,t_2]\psi^D_{j,k,l}[t_1,t_2] \tag{3}$$

$C^D(j,k,l)$ is the output of curvelet transform From theFig. 15 a,b,c are the curvelets at difffernt direction or angle.In the

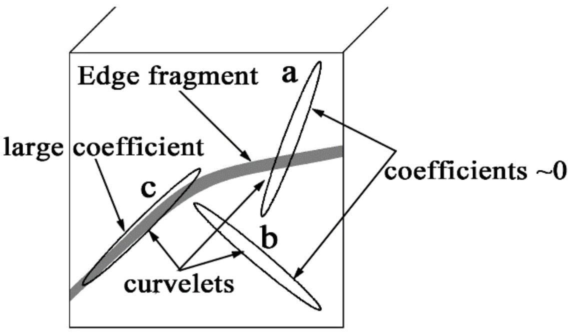
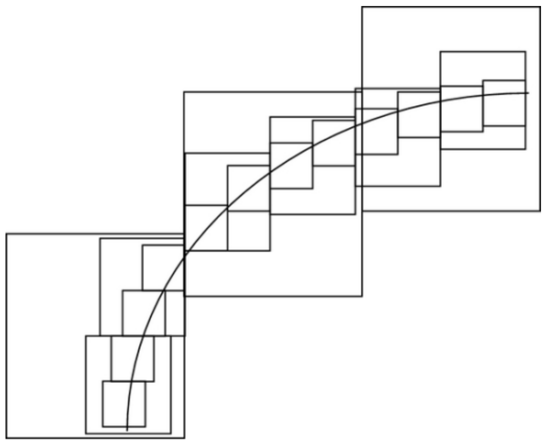


Fig. 15. Working of curvelet transform

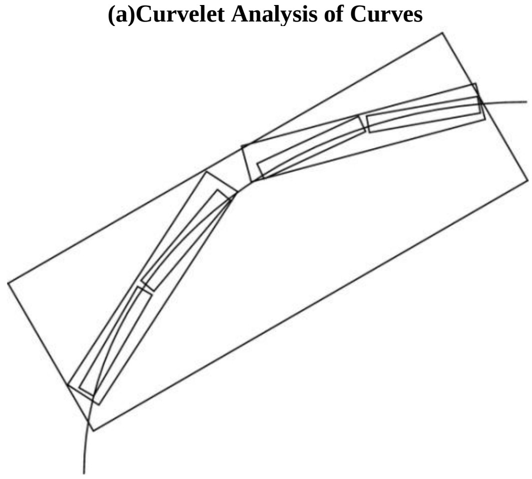
figure the curvelet named c whcih is perfectly alignedwith the curved edge and there for it have large coefficient value. Curvelet a, b will have a low coefficents value close tozero Since they are far from the alignment [14].

5.1 Comparison with wavelet

Fourier series requires a large number of terms to reconstruct a discontinuity good accuracy. Wavelets transform havethe ability to solve this problem of Fourier series, as theyare localized and multiscale. However, though wavelets dowork well in one-dimension, they fail to represent higherdimensional singularities (especially curved singularities). Atthe same time curvelet can model such curved discontinuitieswell. The main idea here is that the edge discontinuity is wellapproximated by curvelet transform and is shown Fig. 16.



(a)Curvelet Analysis of Curves



(b) Wavelet Analysis of Curves

Fig. 16. Comparison between wavelet and curvelet

5.2 Feature extraction using curvelet

Following steps are involved in computing feature vectorusing curvelet transform.

1) Compute Fast discrete cosine transform of the character
2) Extract Coarse Curve Coefficient
3) Find projection of Coarse Curve Coefficient
4) Feature vector is formed by concatenating both projections

5.3 Curve Lab

The software package CurveLab used in this paper forfeature extraction, and is available at http://www.curvelet.org.

6. CHARACTER RECOGNITION USING NEURAL NETWORK

After feature extraction, next step is to character recognition.Character recognition is done using a multi-layer perceptron.Before neural network is used for recognition it has to betrained properly with character database. After neural networkis trained features from the character to be recognised are givento the input of the neural network. From the previous trainingexperience the neural network will classify the character toone of 53 character.

Artificial neural networks (ANN) are the result of academicinvestigations that use mathematical formulations to modelnervous system operations. The aim of neural networks is tomimic the human ability to adapt to changing circumstancesand the current environment. This depends heavily on beingable to learn from events that have happened in the past andto be able to apply this to future situations. In common withbiological neural networks, ANNs can accommodate manyinputs in parallel and encode the information in a distributedfashion.

Artificial neural networks consist of many nodes, i.e. processing units analogous to neurons in the brain. Each nodehas a node function, associated with it which along with aset of local parameters determines the output of the node,given an input. The neural net can generally be a singlelayer or a multilayer net. Fig. 17 shows the block diagaramrepresentation of multilayer ANN [15].

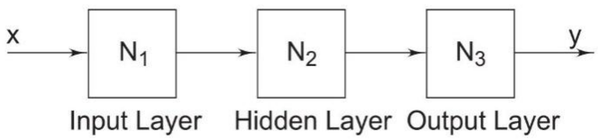


Fig. 17. Block Diagram representation of 3 layered MNN

6.1 Architecture of Neural Network

The architecture of Artificial Neural Network is shown inFig. 18. It consists three layers named input layer, hidden or middle layer and output layer and the neurons are arrangedalong these layers.

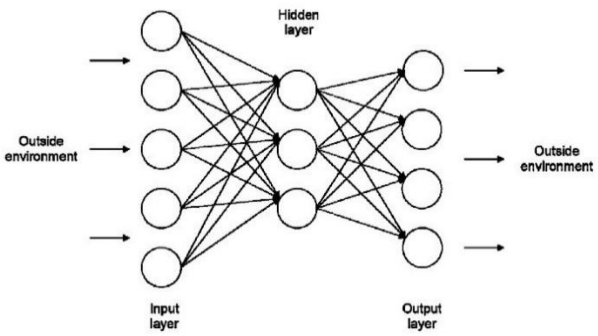


Fig. 18. Feed forward network with 3 layer

It consists three layers named input layer, hidden or middlelayer and output layer and the neurons are arranged alongthese layers. Number of nodes in the input layer depend onfeature vector and in output depend on classes of character.neural network comprises many processing elements knownas neuron and these neurons receives several signals from itsinput links process it and sends to other neurons through outputlinks. The neuron computes the weighted sum of input signals.The mathematic equation for this is given below.

$$X = \sum_{i=1}^{n} x_i w_i \tag{4}$$

Where x is neuron associated weight w.

Then the weighted sum is compares with a threshold value. If the net input is less than the threshold value, the neurondoes not activate and if the net input value is greater than orequal to the threshold, the neuron becomes activated.

$$Y = \begin{cases} +1 \text{ if } X \geq \theta \text{ is even} \\ -1 \text{ if } X < \theta \text{ is even} \end{cases} \tag{5}$$

Here θ is threshold [16]. The above transfer or activationfunction is called sign function. Four common types of activation of neural network used in practical application.Theseare step, sign, linear and sigmoid fuction as shown Fig. 19.

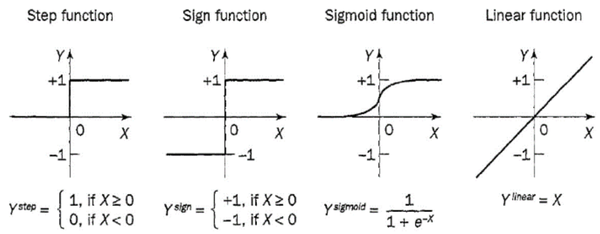


Fig. 19. Activation functions of a neuron

6.2 Back propagation Algorithm

The training algorithm of back propagation involves fourstages. Numbers of different learning algorithms are availableto train the neural network but the most popular one is Backpropagation. Learning in MLP is similar as Perceptron. Theinput patterns are applied to the input layer of the networkand these patterns are processed by input layer and passed tothe next layer. In this way these patterns propagate layer bylayer until the output pattern is generated by output layers.

If the output patters are different from the desired oneerror is calculated and propagates backwards through outputlayer to the input layer. The weights are modified as the errorpropagated.

A neuron determines its output and computes the netweighted input

$$X = \sum_{i=1}^{n} x_i w_i - \theta \tag{6}$$

Where n is number of neurons and is threshold applied toneuron.

$$Y_{sigmoid} = \frac{1}{1 + e^{-X}} \tag{7}$$

The error signal at the output k at iteration p is defined by

$$e_k(p) = t_k(p) - y_k(p) \tag{8}$$

weight in each layer will update based on the error as

$$w_{jk}\ (new) \ = \ w_{jk}\ (old) + w_{jk} \tag{9}$$

As a result of weight update the neural network try toreduce error. Two different training algorithm is used thereare Levenberg-Marquardt backpropagation and Bayesian Regularization backpropagation

6.2.1 Levenberg-Marquardt backpropagation

The Lavenberg-Marquardt algorithm uses an early stoppingcriterion to improve network training speed and efficiency. Todetermine the criterion, all the data are divided into three sets.The first set is the training set for determining the weights andbiases of the network. The second set is the validation set forevaluating the weights and biases and for deciding when tostop training. The validation error normally decreases at thebeginning of the training process. When the network startsto over-fit the data, the validation error begins to increase.The training is stopped when the validation error begins toincrease and the weights and biases will then be derived at theminimum error. The last data set is for validating the weightsand biases to verify the capability of the stopping criterion andto estimate the expected network operation on new data sets.

6.2.2 Bayesian Regularization

Bayesian regularization is a modification of the Levenberg-Marquardt training algorithm to improve the models generalization. Over-fitting problem or poor generalization capabilityhappens when a neural network over learns during a trainingperiod. As a result, such a too well trained model may not perform well on unseen data set due to its lack of generalizationcapability. This approach involves modifying the performancefunction, which is normally chosen to be the sum of squaresof the network errors on the training set (MSE or E_d) [17].

Training stops when any of these conditions occurs:

- The maximum number of epochs (repetitions) is reached.
- The maximum amount of time is exceeded.
- Performance is minimized to the goal.
- The performance gradient falls below min grad. mu exceeds mu max.
- Validation performance has increased more than max failtimes since the last time it decreased (when using validation)

7. RESULTS AND DISCUSSIONS

The experiment is done using Matlab R2013a on a computerhaving Intel i3 processor and 6 GB RAM. In the experimenttrained the network using 40 samples of each 53 classes and atotal of 1760 (40 × 53) samples. The trained network is tested using 10 samples of each classes. Find out system performancechanging the transformation technique in feature extraction.Curvelettransform is used for feature extraction. The performance of the system for these featureswill explain the detail. Extracted feature using these techniquesis used for train neural network. Neural network used arefeed forward net with logsig activation function at hidden andoutput layer as in Fig. 20.

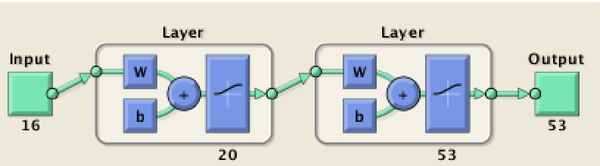


Fig. 20. Artificial Neural network used for recognition

Number of inputs of a neuralnetwork depend on size of feature vector and the number ofnodes at output depends on number of classes.

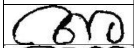
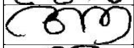
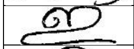
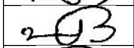
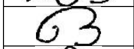
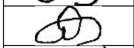
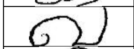
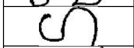
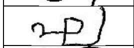
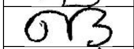
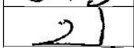
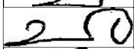
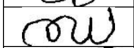
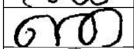
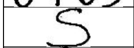
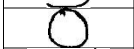
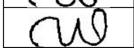
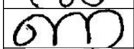
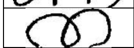
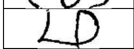
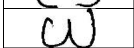
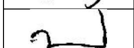
7.1 Performance for Curvelet based feature

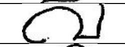
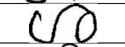
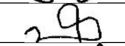
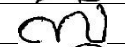
When curvelet transform is used for feature extractionrecognition accuracy of 80.79% is obtained. Since there isno inbuilt toolbox in Matlab to compute curvelet transform. ACurveLab [12] tool box by candes is used.

Feature extraction steps adopted discussed in the earlysection 5.2.

Feature vector of dimension 64 is used for classification.Table I shows result that has obtained.

Table 1Performance of system for different characters using Curvelet feature

Character	FalseNegat iverate	FalsePositi verate	True Positiverat e	TrueNegat iverate
	0.0046	0	1	0.9954
	0	0	1	1
	0.0023	0.3077	0.6923	0.9977
	0.0093	0.25	0.75	0.9907
	0.0023	0.1818	0.8182	0.9977
	0.0069	0	1	0.9931
	0.0047	0.3846	0.6154	0.9953
	0.0069	0.125	0.875	0.9931
	0.0023	0.3571	0.6429	0.9977
	0.0047	0.2	0.8	0.9953
	0.0184	0.6	0.4	0.9816
	0.007	0.4167	0.5833	0.993
	0.0023	0.1	0.9	0.9977
	0.0047	0.3333	0.6667	0.9953
	0.0139	0.5	0.5	0.9861
	0.0092	0.1429	0.8571	0.9908
	0.0023	0.1818	0.8182	0.9977
	0.0047	0.2727	0.7273	0.9953
	0	0.3333	0.6667	1
	0.0047	0.2	0.8	0.9953
	0.0024	0.4	0.6	0.9976
	0.0047	0.2	0.8	0.9953
	0.0092	0.1429	0.8571	0.9908
	0.0024	0.4	0.6	0.9976
	0.0094	0.5385	0.4615	0.9906
	0.0093	0.4	0.6	0.9907
	0.0023	0	1	0.9977
	0.0092	0.1429	0.8571	0.9908
	0.0069	0	1	0.9931
	0.0024	0.4375	0.5625	0.9976
	0.0093	0.3333	0.6667	0.9907
	0.0023	0.25	0.75	0.9977
	0.007	0.3636	0.6364	0.993
	0.007	0.2222	0.7778	0.993

	0.0115	0.2857	0.7143	0.9885
	0.0047	0.4286	0.5714	0.9953
	0.0069	0	1	0.9931
	0.0116	0.375	0.625	0.9884
	0.007	0.2222	0.7778	0.993
	0.0046	0.1111	0.8889	0.9954
	0.0092	0	1	0.9908
	0.0139	0.4286	0.5714	0.9861
	0.0023	0	1	0.9977
	0.007	0.3636	0.6364	0.993

7.2 Graphical User Interface

A graphical user interface created using Matlab GUI creator.Gui created is shown in the Fig. 21 The character to berecognised is loaded by clicking the button load image.

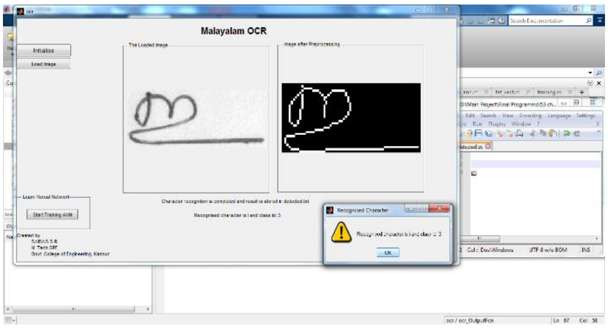


Fig. 21. Graphical User Interface

The image in the left side shows the loaded image and the image on the right side shows a preprocessed image. The detect character will show in file detected.txt file. Also A pop message also appear which contain the id and character recognized is shown in Fig. 22.

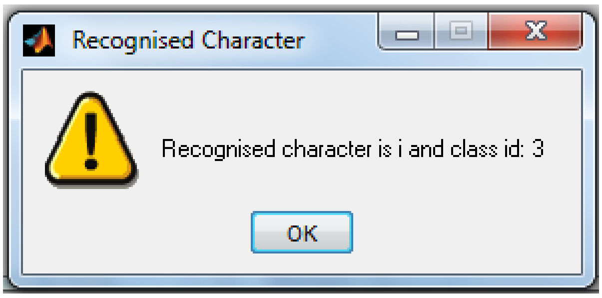


Fig. 22. Detected character in pop-up window

The start training ANN button will start the training ofneural network and attain the weights in each node.

8. CONCLUSION AND FUTURE WORK

Current status of OCR in Malayalam language has beenreviewed. A Malayalam character recognition system is proposed using Curvelet transform and neural network. A graphical user interface is also created for the proposed systemusing Matlab. The proposed system is successful in classifying53 isolated characters. Future work aims at the reduction ofdimension for the feature vector.

9. REFERENCES

[1] Malayalam Script Features [Online]. Available: http://scriptsource.org/scr/Mlym

[2] Malayalam [Online]. Available: http://www.omniglot.com/writing/malayalam.htm

[3] L. Eikvil, "OCR - optical character recognition," 1993.

[4] G. Raju, "Recognition of unconstrained handwritten malayalam characters using zero-crossing of wavelet coefficients," Advanced Computing and ommunications, pp. 217–221, 2006.

[5] B. Philip and R. S. Samuel, "A novel bilingual ocr for printed malayalam-english text based on gabor features and dominant singular values," in Proceedings of the International Conference on Digital Image Processing, 2009, pp. 361–365.

[6] R. John, G. Raju, and D. S. Guru, "1d wavelet transform of projection profiles for isolated handwritten malayalam character recognition," International Conference on Computational Intelligence and Multimedia Applications, pp. 481–485, 2007.

[7] M. A. Rahiman and M. S. Rajasree, "Printed malayalam character recognition using back-propagation neural networks," IEEE International Advance Computing Conference, pp. 197–201, 2009.

[8] B. Philip and R. D. S. Samuel, "An efficient ocr for printed Malayalam text using novel segmentation algorithm and svm classifiers," ACADEMYPUBLISHER, vol. 1, pp. 179–182, May 2009.

[9] A. Rahiman and M. S. Rajasree, "Recognition of handwritten Malayalam characters using vertical & horizontal line positional analyzer algorithm," ICMLC, vol. 4, pp. 404–410, 2011.

[10] A. T. Jia, A. Yahkoob, and S. K, "Malayalam ocr: N-gram approach using svm classifier," International Conference on Advances in Computing,Communications and Informatics (ICACCI), pp. 1799–1803, 2013.

[11] J. John, P. K. V., and K. Balakrishnan, "Offline handwritten Malayalam character recognition based on chain code histogram," pp. 736–741,2011.

[12] [Online]. Available: http://www.curvelet.org/

[13] E. Candes, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelettransforms," MultiscaleModeling& Simulation, vol. 5, no. 3, pp. 861–899, 2006.

[14] T. Guha and Q. J. Wu, "Curvelet based feature extraction," FaceRecognition, pp. 35–46, 2010.

[15] S. Sivanandam and S. Deepa, Introduction to Neural Networks UsingMatlab 6.0, ser. Computer engineering series. Tata McGraw-Hill, 2006.[Online]. Available: http://books.google.co.in/books?id=jTN8RPgyXgC

[16] M. Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems.Addison-Wesley, 2005. [Online]. Available: http://books.google.co.in/books?id=1BxYQnrfv9MC

[17] R. Vilaithong, S. Tenbohlen, and T. Stirl, "Neural network for transformer top-oil temperature prediction."Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.