Conglomeration of Hand Shapes and Texture Information for Recognizing Gestures of Indian Sign Language Using Feed forward Neural Networks

P.V.V.Kishore [#1], S.R.C.Kishore [*2], M.V.D.Prasad [#3]

[#] Department of Electronics and Communications Engineering, K L University
Vaddeswaram, Guntur District, AP, INDIA-522 502
[1] pvvkishore@kluniversity.in
[3] mvd_ece@kluniversity.in
[*] Department of Electronics and Communications Engineering, MESG
Bhoghapuram, Vizinagaram, AP, INDIA-535 216
[2] srckishore@miracleeducationalsociety.com

***Abstract*—This research paper highlights the use of shape and texture information for recognizing gestures of Indian sign language. The proposed method involves extracting the hand segments from the original color gesture images and subjecting them to further processing. In the next stage texture information of the hands in extracted using gabor filter. Again from the segmented hand portions shape is modeled using Chan-Vese(CV) active contour model. Finally both the shape and texture information are merged together to produce a feature vector that essentially represents a sign in Indian Sign Language. To reduce the dimensionality of the feature matrix principle component analysis is applied on the feature matrix. The obtained feature matrix will train a artificial neural network the learns using error back propagation algorithm. Indian sign language database was created for around 36 signs with 10 different signers. For training 4 sets gesture images were used and the remaining 6 sets were used for testing. After extensive testing under various conditions the average recognition rate stands at 98.2%.**

Keyword-Sign Language Recognition, Gabor Filter, Chan-Vese Active Contours, Feed Forward Neural Networks, Recognition Rate.

I. INTRODUCTION

Visual gesture language is employed by a large percentage of deaf populations to communicate among themselves. It has an extensive vocabulary, its own grammatical patterns, and its own rules of usage. For a common person to communicate with a deaf person, a translator is considered necessary to decode sign language into usual spoken language and vice versa**.**

The key element of many sign languages is to recognize hand shapes, signing area, hand movement, and head movement. When two people are observed in conversation of sign language, their bodies bend, facial expressions change constantly, and their hands seem to be constantly in motion. Some signs require no hand movement. These signs are called stationery signs. Some static signs are expressed with one hand and other with two hands. A gesture is a form of non-verbal communication made with a part of the body, used instead of or in combination with verbal communication (Wikipedia).

Sign language recognition aspires to convert sign language into text or speech is an efficient and exact way. This can be accomplished by using digital image processing techniques and neural networks. In this paper are tried to recognize gesture of letters of English alphabets for Indian sign language. Recently sign language gestures recognition has gained a lot of interest by researchers in the areas of computational intelligence, neural networks and image pattern recognition.

Mohammed Mohandas et. Al [1] has proposed an image based systems for Arabic sign language recognition using hidden Markov Model for recognition; they have used a Gaussian skin color model to detect the signal face. The proposed system achieved a recognition accuracy of 98% for a date set of Arabic sign gestures. J.Han, G.Awad et.al.[2] proposed a video based framework for segmenting and tracking skin objects which consists of two parts: a skin color model and a skin object tracking system.Syed Atif Meohadi et. Al [3] examines the possibility of recognizing sign language gestures using sensor gloves. The paper describes the uses of a sensor gloves to capture ASL and uses NN to recognize these gestures. M.Shimada. et.al[4] has proposed an algorithm for figure spelling recognition of static and motion images. The recognition rate is 85.2%.

M.K.Bhuyan et.al [5] developed a framework for hand gesture recognition system based on object based video abstraction technique. The experiment results show that the developed system can recognize signs of

Indian sign language. Rini Akmelia et.al [6] proposed a real time Malaysian sign language translation using color segmentation technique which has a recognition rate of 90%. Nariman Habili et.al [7] proposed a hand and face segmentation technique using color and motion cues for the content based representation of sign language video sequences.

Sign language, similar to spoken language is not confined to a particular region or territory. It is practiced differently around the world. In USA it is known as American Sign Language (ASL) [8],[9] and in Europe[10],[11] British Sign Language (BSL), Africa [12] African Sign Language and in Asia[13] as Arabic Sign Language and Chinese Sign Language[14].

In the last decade there were more efficient algorithms for training[15] and more efficient algorithms for generation [16] with the advancement of powerful computers and bigger parallel corpora.

Unlike America and Europe sign language, India does not have a standard sign language with necessary variations. But in the recent past Ramakrishna missions Vivekananda University, Coimbatore came up with an Indian sign language dictionary. There are around 2037 signs[17] currently available in Indian Sign Language (ISL).

Our previous research[18],[19] show the use of discrete cosine transform and ANN for recognizing gestures with a recognition rate of 93%. The sign language recognition system proposed is signer independent and background independent. Most of our past research [20]-[22] has dealt with videos involving full human head, hands and torso. Using full length human in sign language recognition videos resulted in recognition rates of around 91%. To further increase the recognition rates, the sign images contain only hand shapes.

In this research, a system is developed for recognizing static gestures of Indian sign language. The experiments are conducted on gestures of English alphabets and numbers by using computerized image processing techniques and Neural Networks. The developed system converts gesture signs of Indian sign language into text and voice messages.

The key difficulty in developing such a system is to make the system signer independent. Since different signs have different hand shapes and sizes, our system uses the powerful Gabor transform[23]-[26] and Chan-Vese (CV) active contour models[27]-[29] to extract features from the images along with other image processing techniques. The extracted features from the images are treated with principle component analysis (PCA)[30] to reduce the dimensionality of the feature vector.

Finally the feature matrix forms an input to the neural network[31]-[33] for a parallel distributed processing. The recognized signs are successfully converted into text and voice messages using MATLAB software.

II. System Architecture

The proposed system architecture is exemplified with the help of a block diagram as shown in the figure 1.

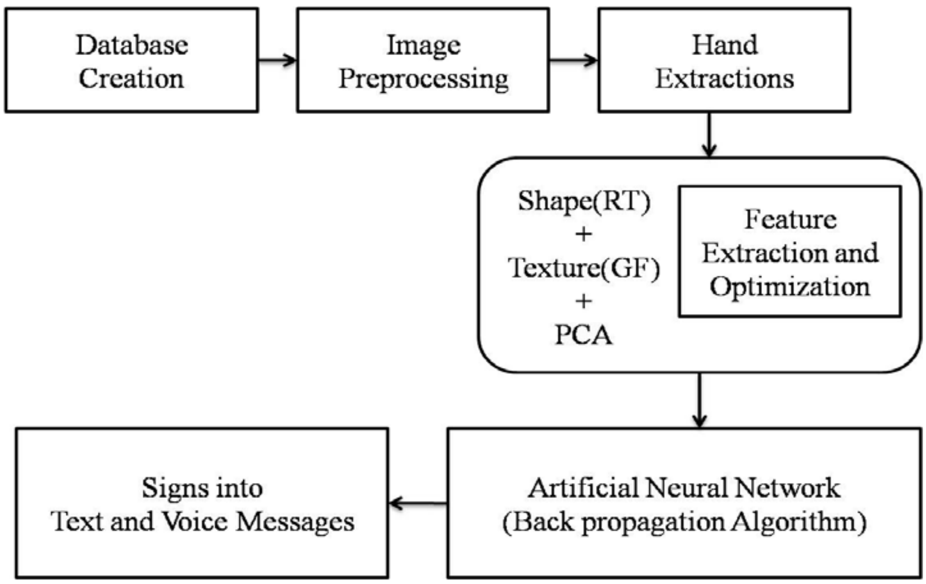


Fig 1. Block Diagram of proposed system architecture

Sign language image database is created by capturing hand images of interpreters of Indian Sign Language from sree santhi ashram school for the deaf, Visakhapatnam. The images are captures using digital camera which creates sign images with a resolution of 1224×1632.The images are preprocessed at various levels to improve the recognition rate. The feature matrix is a mixture of texture features and shape features.

Texture features are extracted using Gabor Filter(GF)[24] and shape features are extracted using Chan-Vese(CV)[27] active contour models. Mixing the features for better classification produces a huge feature vector. This can be reduced in size by applying PCA also increasing the simulation speed.

The feature vector is created for all the 36 signs. Finally a feature matrix is formed by placing the 36 feature vectors in the 36 columns. Each column in the feature matrix represents a unique sign of Indian sign language. This Feature matrix becomes input to a artificial neural network which is a feed forward network. Training of neural network is done with the obtained feature matrix using error back propagation algorithm.

Exclusive testing of the network is performed and the correctly classified signs are converted first into text and then voice messages. Text is converted into audio using simple text to audio conversion algorithm by accessing the Microsoft windows-xp win32 API text to voice program [34].

III. SIGN IMAGE PROCESSING

The function of image pre-processing is to transform the image in such a way that it can make the job of future algorithms simple and effective.

A. Image Resizing

The hand images acquired during image acquisition process have a resolution of 1224×1632. These high resolution images require a large processing time. This can be reduced by using bilinear transformation on pixels of the image without much information loss. Resolution is reduced to 256×256 to facilitate speedy processing in the next stages of processing.

B. RGB to Gray Scale Conversion

RGB images are represented in three color planes. This again creates processing difficulties in the form of increased processing time. This can be controlled by using pixels in any one of the three planes. Since the images are having a red tone, red plane is considered to contain more information.

C. Image Adjustment

The obtained R plane hand image is very low on dynamic range which falls short of distinguishing between different gray level pixels in the image. Dynamic range of the image can be increased by using contrast stretching operation.

D. Image Filtering

Filtering removes noise from the image. This is achieved by averaging pixels in a 3×3 subimage over the entire image. This type of filter is called averaging filter. By applying averaging filter random noise in the image is reduce to great extent.

E. Hand Image Segmentation

Image segmentation process outputs an image separated with similar pixel clusters. In this case these are hands of the signer against the background of the image. The simplest form of segmentation is probably Otsu thresholding which assigns pixels to foreground or background based on grayscale intensity. Figure 2 shows the results of entire pre-processing stages.

IV. TEXTURE FEATURE EXTRACTION USING GABOR FILTER COEFFICIENTS

Gabor filter [25], named after Dennis Gabor, in image processing is used as a Linear Filter for object edge detection. Frequency and orientation representations of Gabor filters are similar to those of the human visual system, and they have been found to be predominantly suitable for texture description and discrimination.

In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. Gabor filters are known more in image processing community than other research areas 2D Gabor filter is an accurate model of a simple cell in the mammalian visual system.

This Filter implements one or multiple convolutions of an input image with a two-dimensional Gabor Kernel. Since features extraction for face recognition using Gabor filters is reported to yield good results.

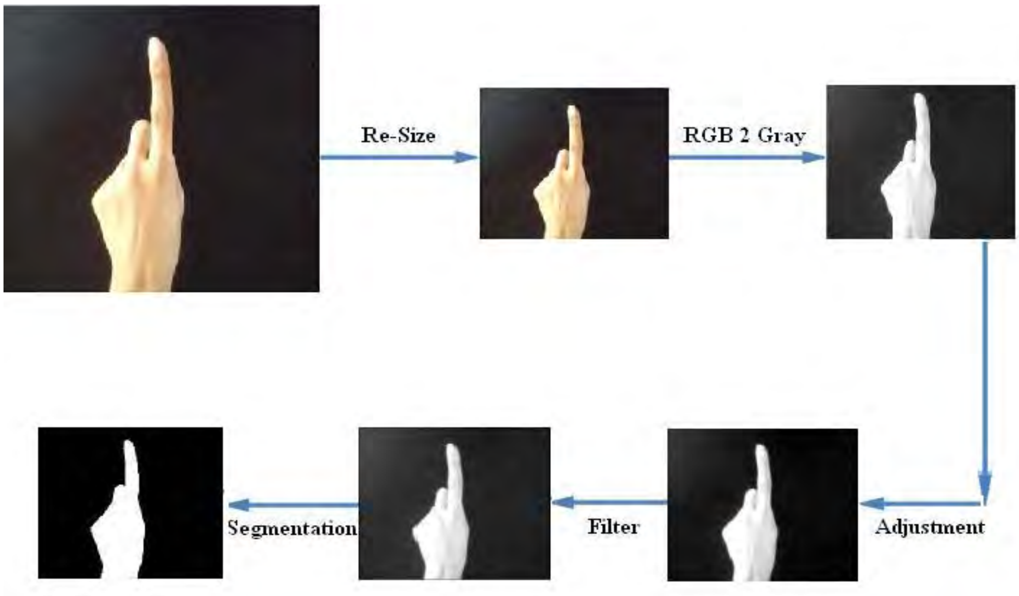


Fig 2: Results from Pre-Processing Stages.

In this paper a Gabor wavelets based feature extraction technique is proposed to model texture. Here we use the following family of two-dimensional Gabor kernels represented mathematically as in eq.1.

$$\left(\begin{array}{l} \vartheta(x, y, \theta, \lambda, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 + y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda} + \psi\right) \\ x' = x\cos(\theta) + y\sin(\theta) \\ y' = -x\sin(\theta) + y\cos(\theta) \end{array} \right) \tag{1}$$

Where (*x, y*) specify the position of a light impulse in the visual field and $\theta, \lambda, \psi, \gamma, \sigma$ are parameters of the wavelet[25].

A. Orientation (θ):

This parameter specifies the orientation of the normal to the parallel stripes of a Gabor function. Its value is specified in degrees. Valid values are real numbers between 0 and 360.

B. Wavelength (λ):

This is the wavelength of the cosine factor of the Gabor filter kernel and herewith the preferred wavelength of this filter. Its value is specified in pixels.

C. Phase offset (ψ):

The phase offset φ in the argument of the cosine factor of the Gabor function is specified in degrees. Valid values are real numbers between 0 and 90.

D. Aspect ratio (γ):

This parameter, called more precisely the spatial aspect ratio, specifies the elasticity of the support of the Gabor function.

E. Gaussian radius (σ):

This is the Gaussian radius of the cosine factor of the Gabor filter kernel and herewith the preferred Gaussian radius of this filter. Its value is specified in pixels.

TABLE I
Shows Parameters Used By Gabor Wavelet

Parameter	Symbol	Values
Orientation	θ	$\left\{0,\frac{\pi}{8},2\frac{\pi}{8},3\frac{\pi}{8},4\frac{\pi}{8},5\frac{\pi}{8},6\frac{\pi}{8},7\frac{\pi}{8}\right\}$
Wavelength	λ	$\left\{4,4\sqrt{2},8,8\sqrt{2},16\right\}$
Phase offset	ψ	$\left\{0,\frac{\pi}{2}\right\}$
Aspect ratio	γ	1
Gaussian radius	σ	$\sigma=\lambda$

A number of Gabor Filters with different scales and orientations are proposed[25]. In this work the Gabor filter used have 5 diverse wavelengths and 8 discrete orientations. The offset phase angles chosen are 0 and $\pi/2$. Thus a total of 40 different Gabor filters are obtained. Real part of Gabor Filters and their magnitudes are represented in Fig 3(a) and 3(b) respectively.

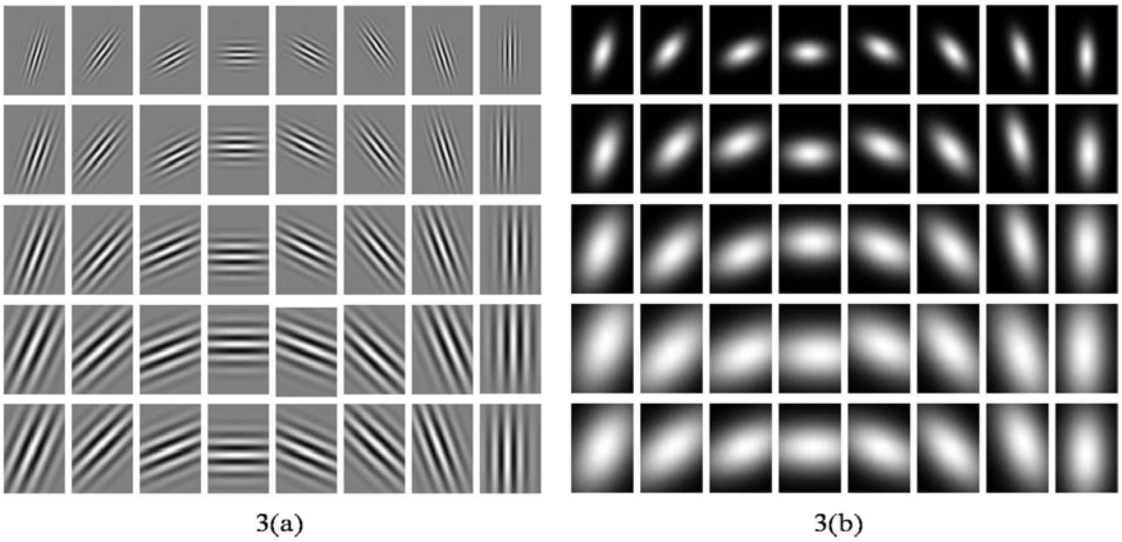


Fig: Gabor Filter 3(a) Real Parts 3(b) Magnitude Response.

The computed Gabor filters are convolved with the sign images $\Im(x,y)$ which results in extraction of texture features at various locations, frequencies and orientations. The convolution process is mathematically modeled [25] using equation 2.

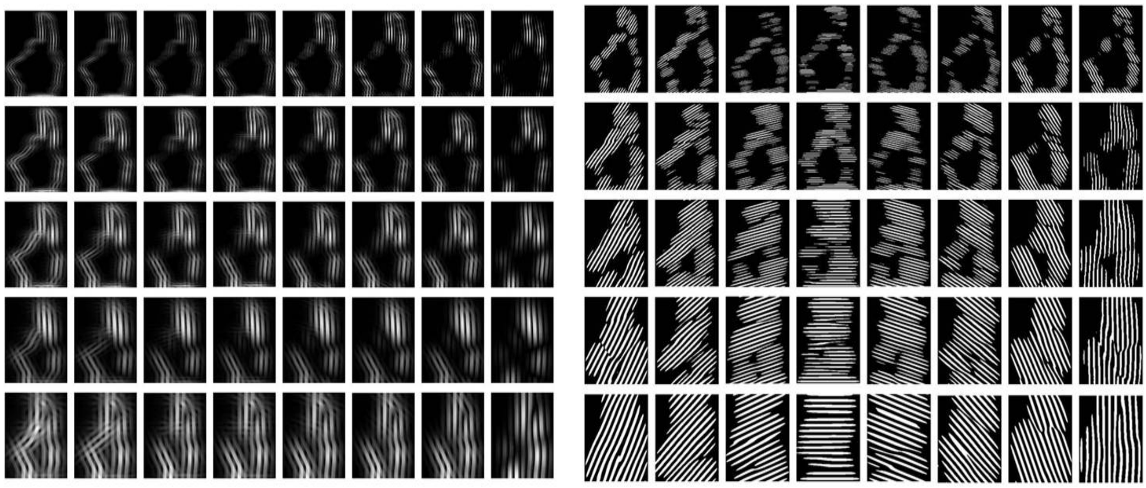
$$G^{(\theta,\lambda,\psi,\sigma,\gamma)}(x,y)=\Im(x,y)*\vartheta(\mathrm{x},\mathrm{y},\theta,\lambda,\psi,\sigma,\gamma) \quad (2)$$

The convolution result for real part and its magnitude are shown in figure 4(a) and 4(b) respectively.

This formulates half of feature matrix containing texture information. The other half of feature matrix consists of shape information extracted using Chen-Vese (CV) active contour models[27]. CV active contour model extracts shape information by using a pre defined initial contour[28].

After multiple iterations the initial contour aligns itself to the boundaries of the objects in an image. The initial contour is either chosen from a standard curve such as circle of can be hand carved from the image plane. Initial contour can be driven towards the edges by using the gradient factor of the image.

Finally shape of the objects in an image are represented by pixels present in the final contour. The next section discusses the issues related to CV model based active contours and their application to gesture shape extraction.



4(a) 4(b)

Fig 4. Convolution Output of Sign Image One with 40 Gabor filters 4(a) Real Parts 4(b) Magnitude Response.

V. Active Contours - Introduction

Active contours or popularly known in the research community as 'snakes' is a active research area with applications to image and video segmentation predominantly to locate object boundaries. They are also used for video object tracking applications.

Active contours come under the category of model based segmentation methods giving good results in the last few years [35]. The active contours was first introduced by Terzopoulos[36].

The basic idea behind active contours is to start with a curve anywhere in the image and move the curve in such way that it sticks to the boundaries of the objects in the image, thus separating the background of the image from its objects.

The original snakes algorithm was prone to topological disturbances and is exceedingly susceptible to initial conditions. However with invention of level sets[37] topological changes in the objects are automatically handled. Nevertheless all active contours are depending on gradient of the image for end the growth of the curve.

Chan and Vese (CV Model)[27] proposed a new level sets method based on Mumford-Shah distance[38] for image segmentation. CV Model for level sets does not necessarily consider gradient for stopping the curve evolution.

A. *Active Contour Models – A theoretical Background*

The active contours are elastic models of continuous, flexible curve that is imposed upon and matched to the image by varying the elastic parameters. The fundamental idea is to make the curve or snake to fit tightly to the boundary of a particular object in the image.

The design of evolution equation is such that the snake can easily embrace the object of importance, to be able to develop a similarity. The first snake model was proposed by Kass[39]. The minimization energy function in order to achieve equilibrium is

$$\mathrm{E}^{Snake} = \int_0^1 \left\{ \mathrm{E}_{int}\left(v(s)\right) + \mathrm{E}_{image}\left(v(s)\right) \right\} ds \tag{3}$$

where the location of the snake on the image is represented parametrically by a planer curve

$$v(s) = \left(x(s), y(s)\right) \tag{4}$$

and E_{int} represents the internal energy of the curve due to bending and E_{image} represents the image forces that push the snake towards the desired object.

The internal energy model was defined as

$$\mathrm{E}_{int} = \frac{\left(\alpha(s)\left|v_s(s)\right|^2 + \beta(s)\left|v_{ss}(s)\right|^2\right)}{2} \quad , s \in [0,1] \tag{5}$$

where $v_s(s)$ First derivative of $v(s)$ and $v_{ss}(s)$ is Second order derivative of $v(s)$ with respect to s. The model of image energy is defined as

$$\mathrm{E}_{image} = -\left|\nabla \Im(x,y)\right|^2 \tag{6}$$

The first derivative of $v(s)$ with respect to $'s'$ gives the rate of change of length of the curve. The coefficient $\alpha(s)$ allows the curve to have smaller or larger degree of contraction of the curve and therefore makes the snake act like an elastic string. The second derivative of $v(s)$ with respect to $'s'$ gives us rate of change of curvature.

The coefficient $\beta(s)$ regulates the rate of the change of the curve in the direction normal to the boundary, preserving the smoothness of the curve. By adjusting these two coefficients, the curve gets an appropriate elasticity and is able to embrace the object of interest.

B. Global Region Based Segmentation- The Chan-Vese Model

The basic idea of Chan-Vese [27] active contour model is to find a contour $\Omega : S \to R^2$, that optimally approximate the image $\Im$ to a single real gray value $\Phi_{internal}$ on the inside of the contour Ω ,and the outside of the contour Ω, by another gray value $\Phi_{external}$.

E^{cv} is the energy function defined by Chan-Vese model as an analogous to piece wise linear Mumford-Shah [39] model which approximates the gray scale image $\Im(x)$ by a piecewise smooth function Ω as a solution to the minimization problem

$$\mathrm{E}^{cv} = \lambda_1 \int_{\Omega} ds + \left[\frac{1}{2} \int_{int(\Omega)} \left(\Im(x) - \Phi_{internal}\right)^2 dx + \frac{1}{2} \int_{int(\Omega)} \left(\Im(x) - \Phi_{external}\right)^2 dx \right] \tag{7}$$

The first term in the eq.5.6 indicates arc length $\arg \min_{\Omega,\Phi} \lambda_1 \times length(\Omega)$ which guarantee evenness of Ω. The second term has two integrals. The first integral function pushes the contour Ω towards the image $\Im$ while the second integral function ensures the differentiability on the contour Ω.

C. The Level Set Model

Sethian and Osher [40] represented boundaries of $\Omega(x)$ implicitly and model their propagation using appropriate partial differential equations. The boundary is given by level sets of a function $\Phi(x)$.

In level sets method, the interface boundary is characterize by a zero level set function $\Phi(x) = 0,\ where\ \Phi : \Re^2 \to \Re, \Omega$. Ω is defined for all values of x,

$$\Omega = \left\{\Phi(x) = 0, x \in \Re^2\right\} \tag{8}$$

The sign of $\Phi(x)$ defines whether x is inside the contour Ω or external to it. The sets $\Omega^{int} = \left\{x, \Phi(x) \le 0\right\}$ and $\Omega^{ext} = \left\{x, \Phi(x) > 0\right\}$. The level set evolves based on the curvature κ of the image objects and assuming the curve moves towards the outward normal $\vec{n}$ defined in terms of parameter Φ as

$$\kappa = \nabla . \left[\frac{\nabla \Phi}{\left|\nabla \Phi\right|}\right] \text{ and } \vec{n} = \frac{\nabla \Phi}{\left|\nabla \Phi\right|} \tag{9}$$

Usually the curve Φ evolution is a time dependent process and the time dependent level set function is represented as $\Phi:\Re^2\times\Re\rightarrow\Re$ as $\Omega(t)=\left\{\Phi(x,t)=0, x\in\Re^2\right\}$. One way to solve is to approximate spatial derivatives of motion and update the position of the curve over time. This method of solving the level sets is prone to unsteadiness due to erroneousness detection of position of the curve.

The sign images are treated with CV active contour level sets as discussed above. The experiments show excellent shape of the hands representing signs as shown in the figure 5.

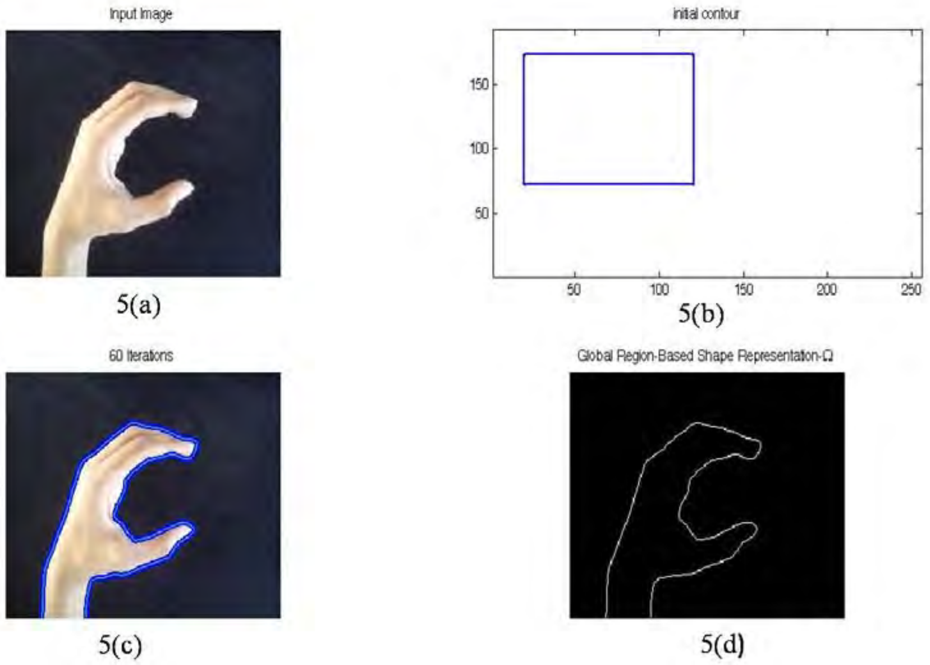


Fig.5(a) Original Image of sign 'C', 5(b) Initial Contour, 5(c) Shows result of CV active contour algorithm , 5(d) Shape Extracted for sign 'C'

Figure 6 shows the shape extraction for somewhat complex sign 'Z'.

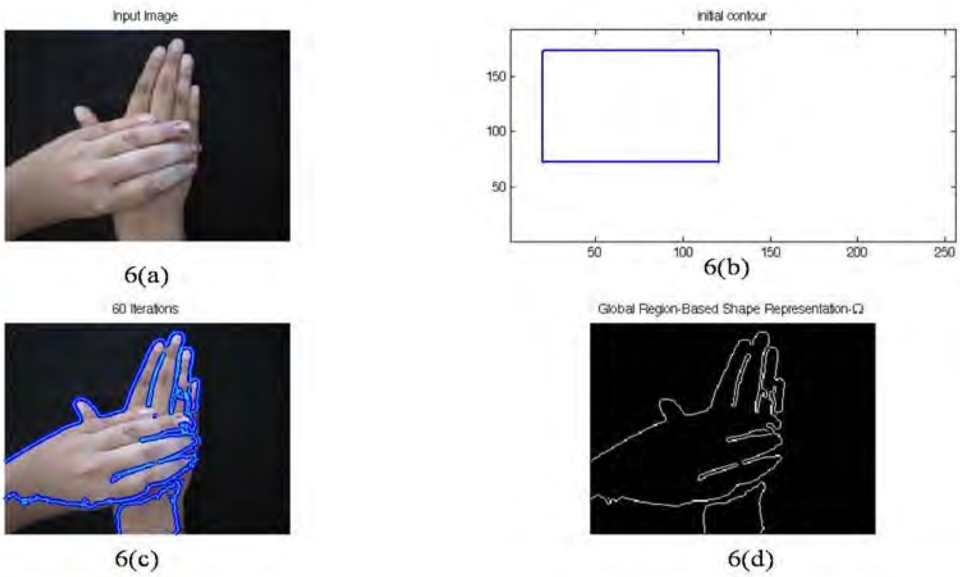


Fig.6(a) Original Image of sign 'Z', 6(b) Initial Contour, 6(c) Shows result of CV active contour algorithm , 6(d) Shape Extracted for Complex sign 'Z'

VI. Feature Vector Creation

The feature vector f^v is a Conglomeration of coefficients of Gabor filter and shape pixels of the final contour obtained from CV algorithm. This combination provides a more perfect feature matrix which uniquely represents signs of indian sign language[17].

40 Gabor filter is applied to single image of resolution 256×256 will provide us with 256×256×40=2621440 coefficients. For all the 36 signs the matrix formed to contain features represents 36×2621440 values. This large data set has to be optimized before applying to the next stage.

Always start with normalizing the input vectors so they have zero mean and unity variance. The mathematics of PCA can be summarized in the following steps. The feature matrix f^v consists of texture numbers from various sign images. Each feature representing a image sign is a 40 Dimension matrix.

The above mentioned issue is addressed by selecting the weights to influence the individual features on the variance in the feature matrix. As a result we are looking for so-called underlying variables that are not directly characterized by the measured features themselves, but rather by a buried combination of them. In other words, the aim is to reduce the dimensionality of our original feature space into a more compact representation, more fitting for our underlying f^v variables.

The initial feature vector for a sign video sequence was 256×256×40. After treating the feature vector with PCA and keeping the threshold above 90%, we create a new feature vector of reduced dimensionality of size 1×256. Finally the entire image for a single sign can be represented with row vector having 256 values.

Finally a complete feature vector can be formed by mixing the 256 values from the shape vector produced using VC active contour model. The above discussed process is presented elaborately in figure 7. The final feature vector for a single sign is represented by 1×512 values. For 36 sample signs the feature matrix consists of 36×512 values.

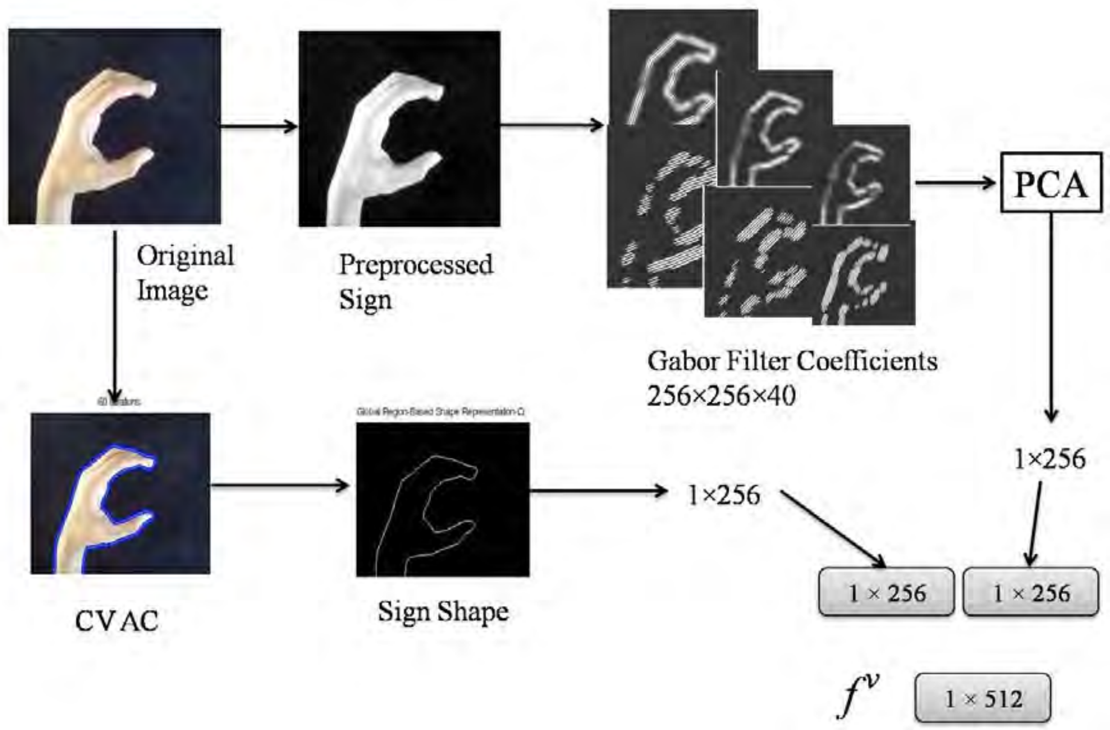


Fig.7 Feature Vector extraction process

VII. Gesture Classification

One of the few systems that can handle our large data matrices is Feed-Forward Artificial Neural Network (ANN)[31]. The dimensionally reduced feature matrix f^v is given as input for training the feed-forward neural network shown in Figure 8.

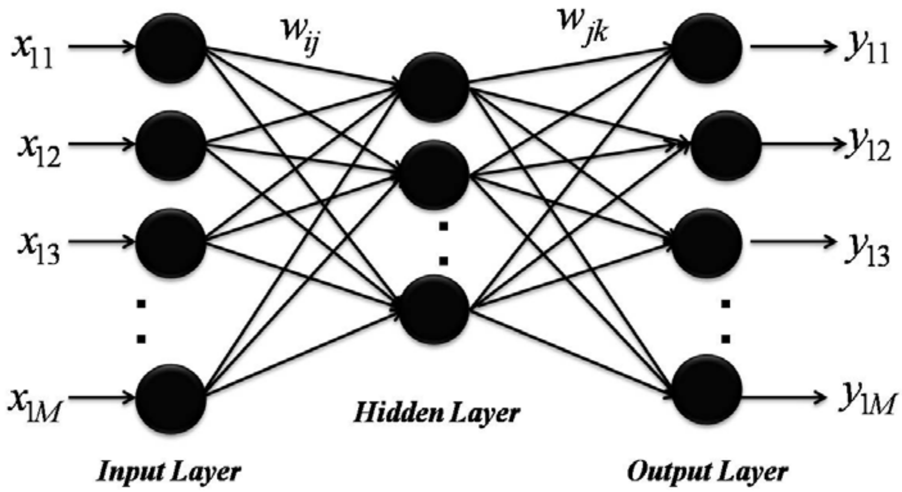


Fig. 8 Feed Forward Neural Network Architecture

Generally, a Feed-Forward neural network is a combination of three layers of neurons: input layer, hidden layer and output layer. The neurons in these layers are activated by using a non linear sigmoid activation function [41].

Let $x_{i,j}(itr)$, where $1 \le i \le N, 1 \le j \le M$, be the input to the neural network derived from feature matrix f^v. Where M and N denote the number of columns and rows of f^v and itr is the number of iterations called Epochs in neural network terminology. The neural network outputs are denoted by $y_{i,j}(itr)$ where $1 \le i \le N, 1 \le j \le M$.

A. Training and Testing the Network

An artificial neural network is employed to accomplish the task of recognizing and classifying gesture signs. Initially the neural network is trained with 36 signs of a particular signer. Consequently the network has 36 neurons in the input layer and 36 neurons in the output layers along with 256 neurons in its hidden layer. This particular neural network object can take 36 input images and classify into 36 signs.

The size of our target matrix is 36×256. Each row in the output matrix is designed to uniquely represent a sign. Out of 256 locations, five locations are filled with ones and remaining are zeros and these locations of one's change with respect to signs. Each row in the target matrix represents a sign. The neural network object created is feed forward back propagation network as shown in Figure 9. The weights and bias for each neuron are initialized randomly and network is ready for training.

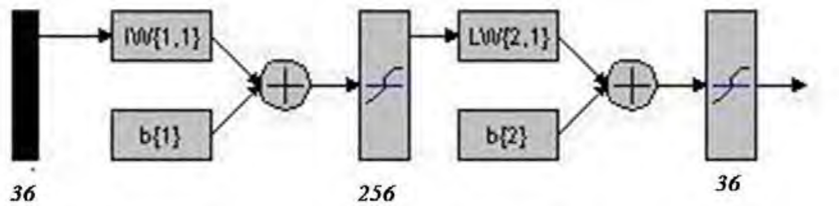


Fig. 9 Neural network architecture for gesture classification trained with one sample per sign

The training process requires a set of examples of proper network behaviour, network inputs and target outputs. During training the weights and biases of the network are iteratively adjusted to minimize the network performance function which in case of feed forward networks is mean square error.

The network is trained with 36 samples for 36 alphanumeric sign images under different conditions. The number of epochs exercised for training is 8695. The system was tested with 36 images previously unseen by the network in the testing phase.

The network was tested more than once during testing phase. The mean squared error tolerance was fixed at 0.0001 for training the samples. The learning rate and momentum factor were chosen as 0.25 and 0.9.

The hidden and output neurons were activated using hyperbolic tangential sigmoid transfer function. The mean square error versus epoch graph is shown in Figure 10.

In the next phase the network is trained with two sets of images i.e.72 alphanumeric images and tested with 72 known and 36 unknown images. The entire testing and characterization is presented in the next section.

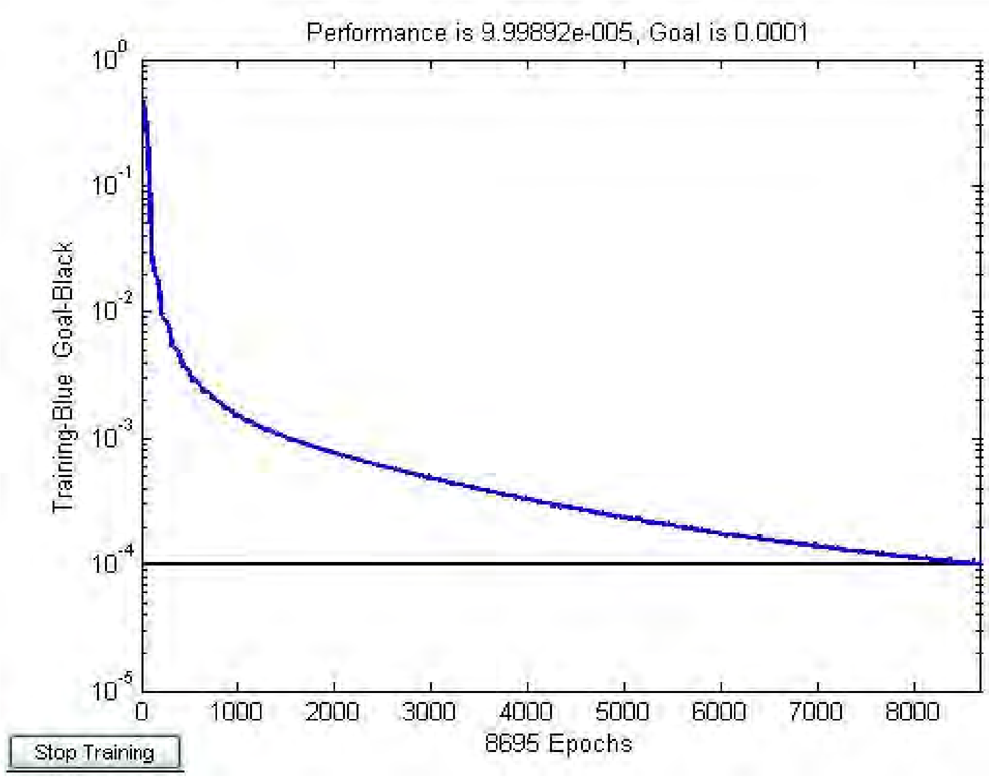


Fig. 10 Mean Square Error versus Epoch Graph

B. Conversion to Voice and Text

Finally the output of neural network is a code representing text corresponding to a particular sign. During testing phase of neural network for a given sign image the network outputs the best match for that particular sign, which is converted to a text message through look up table procedure.

This text output is displayed on the screen. Text is then converted into audio using simple text to audio conversion algorithm by accessing the Microsoft windows-xp win32 API text to voice program[34].

VIII. RESULTS AND DISCUSSION

The evaluation criterion of the sign language interpreter is carried out by calculating the recognition rate formulated according to the following equation below

$$True\ Classification\ Rate(\%) = \mathrm{T} = \frac{Correctly\ Classified\ Signs}{Total\ Number\ of\ Signs\ Used\ for\ Classification} \times 100 \qquad (10)$$

The percentage of recognition is a parameter used by almost all researchers to rate the performance of the sign language interpretation systems developed across the world.

To compute T we used RGB images of Indian sign language [17] alphabets and numbers from 1 to 10, a total of 36 signs with 10 different signers as shown in figure 11. Total number of sign images we have is 360, out of which 144 are used for training and remaining 216 sample sign images are used for testing.

The Table II shows the recognition rates T in percentage for individual alphabets and numbers. Interestingly, this work is compared with the results obtained using a recognition system [42] for American Sign Language (ASL) on sign images using Hough Transform for feature extraction and Backpropagation based artificial neural network.

The table 2 gives the average total recognition rate for the proposed system with sign images close to 98.61% which is better than the result obtained by the authors in [42]. But the only difference in the approach of [42] when compared to the method discussed in this paper is the amount of data samples used for training.

The neural network classifier is trained with 10 samples per sign. The recognition rate improved to 98.61% but with an increase in number of epochs. There is a tradeoff between processing power and recognition rate in case of sign language recognition system.

Fig. 11 Hand Gesture of Indian Sign Language used for our research

TABLE III
RECOGNITION RATES OF V2MI WITH IMAGES COMPARED WITH METHOD IN [42]

Sign	Correctly Recognized Signs for a total of 10 samples/sign	True Classification Rate (%) with Gabor+CV+ANN	Correctly Recognized Samples for a total of 15 samples/sign	Recognition Rate (%) Using Sobel+Hough Transform+ANN [42]
A	10	100	15	100
B	10	100	14	93.33
C	10	100	15	100
D	10	100	15	100
E	9	90	14	93.33
F	10	100	15	100
G	10	100	12	80
H	10	100	10	66.67
I	10	100	12	80
J	10	100	13	86.67
K	10	100	13	86.67
L	10	100	13	86.67
M	9	90	14	93.33
N	9	90	12	80
O	10	100	12	80
P	10	100	12	80

Sign	Correctly Recognized Signs for a total of 10 samples/sign	True Classification Rate (%) with Gabor+CV+ANN	Correctly Recognized Samples for a total of 15 samples/sign	Recognition Rate (%) Using Sobel+Hough Transform+ANN [42]
Q	10	100	13	86.67
R	10	100	13	86.67
S	9	90	12	80
T	10	100	12	80
U	10	100	13	86.67
V	10	100	14	93.33
W	10	100	14	93.33
X	10	100	11	73.33
Y	10	100	11	73.33
Z	9	90	12	80
1	10	100	15	100
2	10	100	15	100
3	10	100	15	100
4	10	100	14	93.33
5	10	100	14	93.33
6	10	100	14	93.33
7	10	100	13	86.67
8	10	100	12	80
9	10	100	12	80
10	10	100	11	73.33
Total	**355**	**98.61**	**471**	**87.22**

The results clearly indicate that the procedure followed for sign language recognition using our proposed method merging the shape and texture information has a clear edge over the procedure discussed in paper[42] using Sobel edge operator, Hough Transform and ANN.

The next table i.e. Table III consolidates all the experiments performed on the network using different training and testing data. The first column gives the number of samples used for training. Number of epochs per training is provided in column two. Column three gives the known samples and unknown samples data set used to test the network.

The fourth column tabulates the results of testing and correctly classified signs. Finally column five formulates the True classification rates with various training and testing data. Last row of the table III indicates average results obtained for the proposed gesture classification system using Gabor transform and CV model with backpropagation based artificial neural networks.

TABLE IIIII
GESTURE CLASSIFICATION RESULTS SUMMARY

Number of samples for training	*Number of epochs for training*	*Number of known and unknown samples for testing*	*Number of Correctly Recognized samples*	*Recognition rate calculated from eq. 10*
36	*8695*	*36*	*33*	*91.67%*
72	*13572*	*72*	*70*	*97.22%*
108	*15304*	*180*	*178*	*98.89%*
144	*18345*	*360*	*358*	*99.44%*
90	***13979***	***162***	***159.75***	***98.61%***

A plot of number of samples used for traing the network against the number of epochs and true classification rate reveals the performance of the system. Figure 12 shows this performance index. The plot reveals a very interesting actuality about artificial neural networks. If the training data is small, the numbers of epochs are small and the classification rate is poor.

Increasing the training samples profoundly increases the total number of epochs which in turn increases the time of training the neural network. But as can be observed the classification rate increased enormously providing excellent performance.

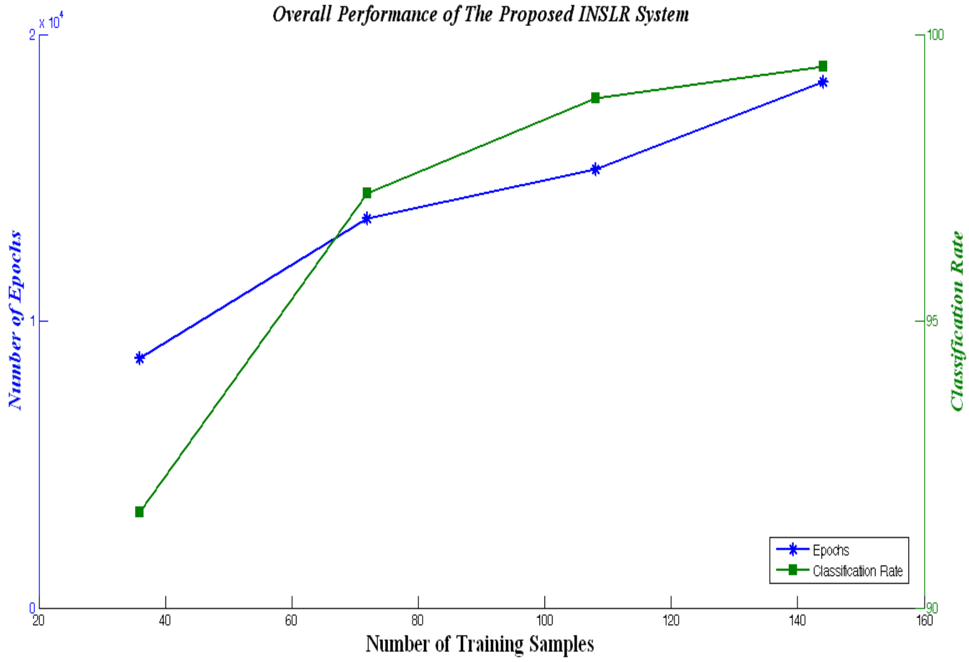


Fig 12. Graph showing performance of proposed recognition system

IX. CONCLUSION

By merging hand textures and hand shapes to design a feature vector that has successfully used to train a neural network the system classification rate has improved promisingly. Gabor filter algorithm models hand textures and CV active contour models hand shapes perfectly. These perfectly extracted hand segments are preserved in texture has a large size. PCA is used to effectively to reduce the size of the feature matrix. The classifier is trained for different inputs combinations of 36 signs from Indian Sign Language with 10 different signers. Testing is also performed with different combination of gestures and the average recognition rate resulted is around 98.61%. It was found that by using neural networks, the time for training is a factor for practical implementation of this system without sacrificing classification rates. The recognition rates produced were excellent when compared with ASL based system in [43] and CSL in [44].

REFERENCES

[1] J. Gauvain, L. Lamel, H. Schwenk, F. Brugnara, R. Schluter, M. Bisani, S. Stuker, T. Schaaf, S. Mohammed, M. Bacchiani, M. Westphal, S.Sivadas, I. Kiss, F. Giron., “Technology and Corpora for Speech to Speech Translation (TC-STAR)”, Technical Report ASR Progress Report, May 2005.

[2] D.M. Gavrila, “The Visual Analysis of Human Movement: A Survey”, Computer Vision and Image Understanding, Vol. 73, No. 1, pp. 82–98, February 1999.

[3] Syed AtifMehdi, Yasir Niaz Khan, sign language recognition using sensor gloves, Proceedings of the 9th International Conference on Neural Information Processing (ICONIP‘02), Vol. 5, 18-22Nov., pp2204-2206.

[4] M. Shimada, S. Iwasaki , and T. Asakura, Finger Spelling Recognition using Neural Network with Pattern Recognition Model, SICE Annual Conference in Fukui, August 4-6.2003.

[5] M.K.Bhuyan and P.K.Bora, “A Frame Work of Hand Gesture Recognition with Applications to Sigb Language”, Annual India Conference, IEEE, pp1-6.

[6] Rini Akmeliawati, Melanie Po-Leen Ooi and Ye Chow Kuang , ‘Real-Time Malaysian Sign Language Translation Using Colour Segmentation and Neural Network’, IEEE on Instrumentation and Measurement Technology Conference Proceeding,Warsaw, Poland 2006 ,pp. 1-6.

[7] Nariman Habili, Cheng Chew Lim and Alireza Moini , ‘Segmentation Of The Face And Hands In Sign Language Video Sequences Using Color And Motion Cues’, IEEE Transactions on Circuits and Systems For Video Technology 2004 , Vol. 14, No. 8, , pp.1086 – 1097

[8] ASL corpus: <http://www.bu.edu/asllrp/>.

[9] Christopoulos, C., Bonvillian, J., 1985. Sign language. Journal of Communication Disorders 18, 1–20.

[10] Atherton, M., 1999. Welsh today BSL tomorrow. In: Deaf Worlds 15 (1),pp. 11–15.

[11] Engberg-Pedersen, E., 2003. From pointing to reference and predication: pointing signs, eye gaze, and head and body orientation in Danish Sign Language. In: Kita, Sotaro (Ed.), Pointing: Where Language, Culture, and Cognition Meet. Erlbaum, Mahwah, NJ,pp. 269–292.

[12] Nyst, V., 2004. Verbs of motion in Adamorobe Sign Language. Poster. In: TISLR 8 Barcelona, September 30–October 2. Programme and Abstracts. (Inter Nat. Conf. on Theoretical Issues in Sign Language Research; 8), pp. 127–129.
[13] Abdel-Fattah, M.A., 2005. Arabic sign language: a perspective. Journal of Deaf Studies and Deaf Education 10 (2), 212–221.
[14] Masataka, N. et al., 2006. Neural correlates for numerical processing in the manual mode. Journal of Deaf Studies and Deaf Education 11 (2), 144–152.
[15] Och, J., Ney, H., 2003. A systematic comparison of various alignment models. Computational Linguistics 29 (1), 19–51.
[16] Koehn, P., 2004. Pharaoh: a beam search decoder for phrase-based statistical machine translation models. AMTA.
[17] Indian Sign language, empowering the deaf, <http://www.deafsigns.org>.
[18] P.V.V.Kishore, P.Rajesh Kumar, "Video based Sign Language Recognition Using DCT and Neural Networks" proceedings at National conference on Recent trends in DSP, Embedded Systems, VLSI, Image Processing, Communications and Electronics Engineering (DEVICE-2010)13-14, November, 2010 organized by ANITS & IEEE Hyderabad, AICTE India,pp71-75.
[19] P.V.V.Kishore, P.Rajesh Kumar, "A static sign language recognition using discrete cosine transform and Ann Back propagation" Elixir International Journal of Advanced Engineering Informatics, 33(2011),ISSN:2229-712x, pp2308-2311.
[20] P.V.V.Kishore, P.Rajesh Kumar, A.Arjuna Rao "Static Video Based Visual-Verbal Exemplar for Recognizing Gestures of Indian Sign Language" CiiT International Journal of Digital Image Processing,Vol(3),Issue(9), Print: ISSN 0974 – 9691 & Online: ISSN 0974 – 9586, pp530-538. DOI: DIP052011009.
[21] P.V.V.Kishore, P.Rajesh Kumar, "A Static Video Based Gesture Recognition for Sign Language using Wavelet Transform and Neural Networks" Proceedings of 2011 International conference on digital convergence organized by International Association of Computer Science and Information Technology (IACSIT) and IEEE, at R.M.K. Engineering College, Chennai, 19-20,Feburary,2011, ISBN:978-1-4244-9325-8,pp105-110.
[22] P.V.V.Kishore, P.Rajesh Kumar, E.Kiran Kumar, S.R.C.Kishore "Video Audio Interface for Recognizing Gestures of Indian Sign Language" International Journal of Image Processing (IJIP), CSC Journals, Kualalumpur, Malaysia, Volume (5) : Issue (4) : 2011,ISSN:1985-2304,pp479-503.
[23] Jain, Anil K., and Farshid Farrokhnia. "Unsupervised texture segmentation using Gabor filters." *Pattern recognition* 24, no. 12 (1991): 1167-1186.
[24] Weldon, Thomas P., William E. Higgins, and Dennis F. Dunn. "Efficient Gabor filter design for texture segmentation." *Pattern Recognition* 29.12 (1996): 2005-2015.
[25] Yang, J., Liu, L., Jiang, T., & Fan, Y. (2003). A modified Gabor filter design method for fingerprint image enhancement. *Pattern Recognition Letters*, *24*(12), 1805-1817.
[26] Bianconi, Francesco, and Antonio Fernández. "Evaluation of the effects of Gabor filter parameters on texture classification." *Pattern Recognition* 40.12 (2007): 3325-3335.
[27] Chan, Tony F., B. Yezrielev Sandberg, and Luminita A. Vese. "Active contours without edges for vector-valued images." *Journal of Visual Communication and Image Representation* 11.2 (2000): 130-141.
[28] Sandberg, Berta, Tony Chan, and Luminita Vese. "A level-set and gabor-based active contour algorithm for segmenting textured images." *UCLA Department of Mathematics CAM report*. 2002.
[29] Osher, Stanley, and Ronald P. Fedkiw. "Level set methods: an overview and some recent results." *Journal of Computational physics* 169.2 (2001): 463-502.
[30] Roweis, Sam T., and Lawrence K. Saul. "Nonlinear dimensionality reduction by locally linear embedding." *Science* 290.5500 (2000): 2323-2326.
[31] Pao, Yohhan. "Adaptive pattern recognition and neural networks." (1989).
[32] LeCun, Yann, Bernhard Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne Hubbard, and Lawrence D. Jackel. "Backpropagation applied to handwritten zip code recognition." *Neural computation* 1, no. 4 (1989): 541-551.
[33] Berry, Michael W., Susan T. Dumais, and Gavin W. O'Brien. "Using linear algebra for intelligent information retrieval." *SIAM review* 37, no. 4 (1995): 573-595.
[34] Microsoft win32 Speech Application Programming Interface (SAPI 5.3),system development, Microsoft Corporation Inc. MSDN.
[35] Ginmo Chung and Luminita Vese. (2003) "Image segmentation using a multilayer level set approach.Technical Report" 03-53, UCLA.
[36] D. Terzopoulos and K. Fleischer (1988) "*Deformable models*". The Visual Computer, Vol.4 No.6, pp306-331.
[37] Luminita Vese and Tony Chan.(2002) "*A multiphase level set framework for image segmentation using the mumford and shah model*". International Journal of Computer Vision, Vol.50,No. 3 pp271-293.
[38] D. Mumford and J. Shah.(1989) "Optimal approximation by piecewise smooth functions and associated variational problems" Comm. Pure Appl. Math, Vol.42,pp577-685.
[39] M.Kass, A Witkin, D Terzopoulos ,"Snakes:active Contour Models", *Int. J. of Computer Vision*, 1987,pp 321-331.
[40] Osher, S., Sethian, J. (1988) Fronts propagating with curvature-dependant speed: algorithms based on Hammilton-Jacobi formulations.J. Comput. Phys. Vol.79 No.1, pp12–49 .
[41] S.N. Sivanandam, M. Paulraj An Introduction to Neural Networks, 2003, Vikhas Publications Company Ltd.India.
[42] Qutaishat Munib et.at., (2007). American Sign Language (ASL) recognition based on Hough Transform and Neural Networks, Expert Systems with Applications 32, Science Direct, Elsevier ltd., pp24-37. DOI: 10.1016/j.eswa.2005.11.018
[43] Paulraj M P,Sazali Yaacob, Hazry Desa Hema, C.R.Wan, Mohd Ridzuan, Wan Ab Majid, "Extraction of Head and Hand Gesture Features for Recognition of Sign Language" 2008 IEEE International Conference on Electronic Design, December 1-3, 2008, Penang, Malaysia.
[44] G. Fang, W. Gao, X. Chen, C. Wang, and J. Ma, "Signer-Independent Continuous Sign Language Recognition Based on SRN/HMM," Proc. Gesture Workshop, pp. 76-85, 2001.