Journal of Theoretical and Applied Information Technology
31st July 2016. Vol.89. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

INDIAN SIGN LANGUAGE RECOGNITION: A COMPARISON BETWEEN ANN AND FIS

D.ANIL KUMAR, P.V.V.KISHORE, N.VENKATRAM, B.LEELA RANI, A.HASITHA, T.V.N.S.H.SABARISH, R.SAI CHANDRA

Department of Electronics and Communications Engineering, K.L.University, Guntur.

E-mail: danilmurali@gmail.com, pvvkishore@kluniversity.in, venkatram@kluniversity.in, bevaraleelarani@gmail.com, hasithaamruthavakkula@gmail.com, sabarishtavva123@gmail.com, saichandraravipati@gmail.com

ABSTRACT

The focus of this paper is to compare two artificial intelligent paradigms for gesture recognition from videos of Indian sign language with artless backgrounds. Hand and Head segmentation giving rise to shape features for the entire video sequence is inputted to Artificial Neural Networks (ANN) and Fuzzy Inference Engine (FIS). Chan Vase active contour extracts shape models for the classifier. Two classifiers are inputted with same set of training and testing samples. The classifiers are compared based on their data handling measured using recognition rate and execution times measured by training and testing times. The results indicate a tradeoff between training data size and execution times of ANN and FIS. The classifiers were tested on 86 gestures from 10 different signers.

Keywords: *Active Contour Shape analysis, Continuous sign language, Hybrid feature vector, Fuzzy Inference Engine (FIS,) Artificial Neural Network (ANN).*

1. INTRODUCTION

Sign language is a computer vision based language for deaf and hearing people, which involves using hands, face and body. Present work complex in continuous sign language video sequences problem in computer vision. Many works focus in the signer dependent sign language recognition. The following system involves preprocessing, segmentation, feature extraction and recognition systems. The important parameters are hand movements and head location in video frames. This work compares pictures, distinct and continuous Indian sign languages videos under normal patterns with easy backgrounds by using ANN and FIS.

The vision communication system uses hand shapes, hand movements, hand orientation, body movements, facial expressions and head movement in speech communication. Hence, a mute person has to depend on visual sense to a large extent and any learning and communication aids will help them learn faster and communicate better. Usually human interpreter trained in sign language understanding is used as a bridge between the normal people (with hearing sense) and mute persons (without or with low hearing sense).

Empowerment of the hearing impaired can be achieved using a mobile based application that can understand sign language and translate it to speech and vice versa. The solution proposed is to develop a machine interpreter that can facilitate deaf people to articulate themselves at any location in the presence of people with hearing sense.

This machine translation of sign language acts an interpreter between these two groups of humans i.e. with and without hearing sense, intendeds to replace the human interpreter with machine interpreter. Sign language recognition is a major research area that encompasses video image analysis, shape extraction, feature optimization and pattern classification working in tandem to convert video signs into text and voice messages.

Previous research in the area show a variety of methods applied to achieve this objective and to a certain extent achieved by most of the researchers.

Sign language recognition acts as a machine interpreter (MI) between a mute person and normal person. Active contours energy function is formulated by amalgamating energy function from boundary and shape prior elements. Artificial Neural Network is constructed to classify and recognize gestures from video frames of signers. The proposed VVMI [1] for SLR offers a recognition rate of around 93%.

The proposed method involves extracting the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

hand gestures form original color images. The segmented hand positions shape modulated using Chan-Vese (CV) active contour model. The shape and texture information are merged to produce a feature vectors will train an artificial neural network [2] by using error back propagation algorithm. After extensive testing the average recognition rate stands at 98.2%.

Kishore Pvv et al [3] proposed 4-Camera model for segmented hand gestures, shape feature extracted with elliptical Fourier descriptors and pattern classification using artificial neural network with back propagation algorithm. The average recognition rate for 4 Camera model is around 95%.

Wang hanjie et al [4] has developed sparse observation alignment for sign language recognition. Sparse observation is used to characterize each sign with typical hand posters. The proposed method is compared with HMM and DTW methods the accuracy speed is higher than 1/10 time.

Tripathi kumud et al [5] proposed continuous Indian sign language recognition. He solved this problem by using gradient based key frame extraction method. After extraction he used orientation histogram with PCA for reducing dimension features.

A vision based approach to classify facial gestures [6] for communication by using artificial neural network (ANN) to recognize lip movement. The results average recognition rate stands at 88%.

Ahmad Zaki Shukor et al [7] proposed Glove approach for Malaysian sign language detection. In this paper first to recognize gestures and after assembling the data glove in translating likes alphabets, numbers and words. The resultant average accuracy of glove data for translating all gestures is around 89%.

Caibo et al [8] proposed active contour segmentation models based on automatic active contour. In this method used to avoid the time taken process of human choosing active contour and increase the accuracy of segmentation.

G. Fang et al [9] has developed signer independent continuous sign language recognition HMM, Self-Organizing Maps and Recurrent Neural Networks. They used 208 signs form Chinese sign language and reported a recognition rate of around 92%.

HMMs compensate well for time and amplitude variances for speech and character recognition [10]. This property makes HMMs a perfect method for hand gesture recognition [11–13]. The expense for efficiency in this case comes from collection of huge data and considerable time required to estimate matching parameters in HMMs.

One of the foreseeable difficulties for using artificial neural networks is that the information is bounded by weight matrix. Conventional non-fuzzy or crisp classification techniques strictly induce a sharp cut between class labels, which is not always reliable for real time engineering applications. On the other hand a class distributed membership is assigned in fuzzy classification algorithms [14].

The partitions between fuzzy classes are “soft.” Since 1965, efforts are on to design algorithms for fuzzy classification and apply the same to pattern recognition and decision systems. M.C.su projected Hyper rectangular Composite NNs in [15], for hand shape classification using a fuzzy rule base approach.

In an adaptive fuzzy expert system by Holden and Owens [13], signs were categorized on start and end hand shapes with finger motion, using triangular fuzzy membership functions, whose parameters were found from training data.

This research proposes a sign language recognition system with shape analysis that builds the feature vector for the classifier. Here we compare active contour based segmentation with Fuzzy Inference Engine and Artificial Neural Network which is trained Least Squares algorithms.

2. LEVEL SETS-REVIEW

Active contours are a model based image segmentation algorithms built on the shoulders of total variational methods [16, 17]. Variational methods define a solution space for the problem and builds a mathematical model that becomes linear during optimization process. First models were introduced by Terzopoulous [18,19]. An initial smooth contour that deforms itself actively towards object edges ensuing in a solution space consisting of object boundaries in the image.

Two initial conditions while defining the snake active contour model are, the solution space image should be very much similar to original image and they should also exhibit spatial smoothness. For certain class of images this works extremely well. But as the problem domain increases the snakes model gives unstable solutions for small changes in pixel values. The stability of the active contours is increased by using the concepts of level sets [20], which can handle object deformities automatically.

Most of the active contours end their growth based on image gradients. Chan and Vese (CV Model) [21] model uses level sets and the growth of the curve is controlled by Mumford-Shah distance [22]. CV Model for level sets does not consider gradient for stopping the curve evolution.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The active contours are elastic models of unbroken, stretchy curve that is levied upon and matched to the image objects by fluctuating the stretchy parameters. The fundamental idea is to make the curve or snake to fit tightly to the borders of a particular image object.

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 [23]. The minimization energy function in order to achieve equilibrium is

$$E^{Snake} = \int_0^1 E^{Snake}(v(s))ds = \int_0^1 E^{\text{int}ernal}(v(s))ds + E^{image}(v(s))ds + E^{CoN}(v(s))ds \quad (1)$$

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

$$\chi(s) = (x(s), y(s)) \quad (2)$$

and $E^{\text{int}ernal}$ 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. E^{CoN} is the constraint that helps keep the snake movements smooth in all directions.

The internal energy model was defined as

$$E^{\text{int}ernal} = \frac{\left(\alpha(s)\left|\chi'(s)\right|^2 + \beta(s)\left|\chi''(s)\right|^2\right)}{2} \quad (3)$$

Where $\chi'(s)$ First derivative of $\chi(s)$ which tracks changing curve length and $\alpha(s)$ maintains the degree of contraction in all directions. Similarly, $\chi''(s)$ is Second order derivative of $\chi(s)$ with respect to s representing changes in snake curvature and $\beta(s)$ normalizes curvature movements in the direction of the normal along the snake boundary. The model of image energy is defined as

$$E^{image} = \left|\nabla f(x,y)\right|^2 \quad (4)$$

This model is further refined by Chan-Vese[21] which finds a contour $\hat{\square}: s \to \Re^2$, that approximates the object regions in image $f(x,y)$ into a single real gray value $\square_{\text{int}ernal}$ internal to boundary of the contour $\hat{\square}$ and $\square_{external}$ to exterior of the boundary $\hat{\square}$. Energy function in CV model is represented with linear Mumford-Shah[22] model which approximates a 2D function $f(x,y)$ by a piece wise smooth function $\hat{\square}$ giving rise to distance minimization problem defined as

$$E^{chan-vese}(\hat{\square}, \square^{(I)}, \square^{(E)}) = \min_{\Theta, \square^{(I)}, \square^{(E)}} \chi_2 \left[\iint_{\text{int}(\Theta)} (I^{xy} - \square^{(I)})^2 \hbar\left(\hat{\square}^{xy}\right) + \iint_{ext(\Theta)} (I^{xy} - \square^{(E)})^2 (1 - \hbar(\hat{\square}^{xy}))dxdy \right] + \chi_1 \int_{\Theta} \left|\nabla \hbar(\hat{\square}^{xy})\right| dxdy \quad (5)$$

The last term in the eq.17 indicates arc length which guarantee evenness of $\hat{\square}$ .The first term has two integrals. The first integral function pushes the contour $\hat{\square}$ towards the image $f(x,y)$ while the second integral function ensures the differentiability on the contour $\hat{\square}$. χ_2 and χ_1 are the regularization parameters which define the percentage of smoothness required for a particular set of pixels.

Sesthian and Osher [24] represented boundaries of $\hat{\square}(x,y)$ implicitly and a set of partial differential equations model their propagation around the edges in the image. Initial level set function $\phi(x)$ is the boundary trace. The interface boundary in the level set model is parametrized by a zero level set function $\phi(x) = 0$ where, $\phi: \Re^2 \to \Re$. $\hat{\square}$ is defined for all values of x.

$$\hat{\square} = \left\{\phi(x) = 0, x \in \Re^2\right\} \quad (6)$$

The sign of $\phi(x)$ controls the pixel x as it is inside the contour $\hat{\square}$ or external to it. The sets $\square_{\text{int}ernal} = \{x, \phi(x) \le 0\}$ and $\square_{external} = \{x, \phi(x) > 0\}$. The curvature κ wheels the level set towards the image objects and the curve smoothness is from the outward normal $\vec{n}$ in terms of parameter ϕ as

$$\kappa = \nabla.\left[\frac{\nabla\phi}{|\nabla\phi|}\right] \text{ and } \vec{n} = \frac{\nabla\phi}{|\nabla\phi|} \quad (7)$$

Here the curve $\hat{\square}$ evolution is a time dependent process and the time dependent level set function is represented as $\phi: \Re^2 \times \Re \to \Re$, $\square(t) = \left\{\phi(x,t) = 0, x \in \Re^2\right\}$. One way to solve is to approximate spatial derivatives of motion and update the position of the curve over time.

Journal of Theoretical and Applied Information Technology
31st July 2016. Vol.89. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3. ARTIFICIAL NEURAL NETWORKS

An artificial neural network is employed to accomplish the task of recognizing and classifying gesture signs. The neural network has 130 neurons in the input and 36 neurons in the output layers along with 256 neurons in its hidden layer. This particular neural network object can take 130 input images and classify into 36 signs.

The size of our target matrix is.36X196 Each row in the target matrix represents a sign. The neural network object created is feed forward back propagation network as shown in Figure 1. The weights and bias for each neuron are initialized randomly and network is ready for training.

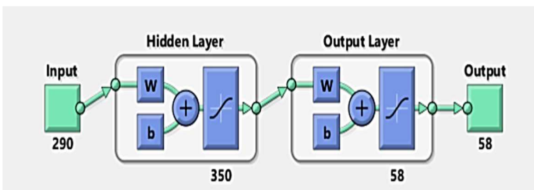

Figure.1 Neural network architecture for gesture classification

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 130 samples for 36 alphanumeric sign images under different conditions. The number of epochs exercised for training is 15000. The system was tested with 194 images previously unseen by the network in the testing phase.

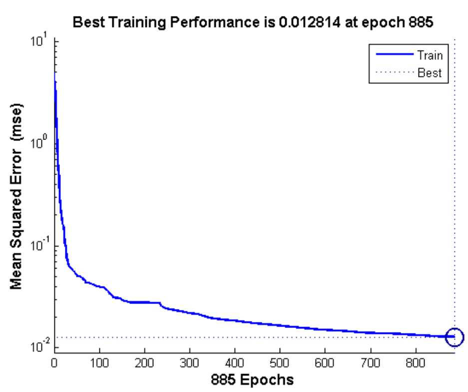

Figure 2.Mean Square Error versus Epoch Graph

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 2.

4. FUZZY INFERENCE SYSTEMS

The fuzzy inference system (FIS) used for gesture classification in this work is the most basic design used by many pattern classification systems [25]-[28]. A FIS is designed for recognizing signs from the hybrid feature matrix that represents shapes and position vectors of hands and head portions in continuous sign videos.

The FIS used for classification of gestures is designed to be a basic one which resembles to that of a traditional FIS proposed by Takagi, Sugeno and Kang[29].

The problem of fuzzy rule base classifier is to find an optimal mapping $\vec{\mathbf{M}}$ from the feature matrix $\vec{\boldsymbol{F}}$ into the decision vector $\vec{\boldsymbol{D}}$ given by $\vec{\boldsymbol{M}}:\vec{\boldsymbol{F}}\rightarrow\vec{\boldsymbol{D}}$.

Suppose that a pattern $\vec{\boldsymbol{F}}$ is represented in terms of N features $\{\vec{F}_1,\vec{F}_2,.....,\vec{F}_N\}$.During the classification process, any extracted feature matrix $\vec{\boldsymbol{F}}$ is to be assigned into one of the **P** possible classes of gestures $\{C^1,C^2,.....,C^P\}$, based on its feature values. Thus the classification process is a mapping from $\vec{\boldsymbol{F}}$ to $[0,1]^P$.

If the gesture classes are fuzzy classes, then the scheme converts into a fuzzy classification system. The classification method is a mapping from $\vec{\boldsymbol{F}}$ to $[0,1]^P$. So a fuzzy rule-based scheme implements a gesture classification system.

The inference rules used are in the form of

$$\begin{aligned} &rule^i : IF\ \mathrm{F}_{x1}^{vect}\ is\ \mathrm{F}^{1i}\ and\ \mathrm{F}_{x2}^{vect}\ is\ \mathrm{F}^{2i}......,\\ &and\ \mathrm{F}_{xN}^{vect}\ is\ \mathrm{F}^{Ni}\ THEN\ y\ is\ C^{ip} \end{aligned} \tag{8}$$

Where $\boldsymbol{F}^{ij}, \boldsymbol{j}=[1,\boldsymbol{N}]$ are fuzzy set and $\boldsymbol{y}^{iP}, \boldsymbol{i}=[1,\boldsymbol{M}]$ are real numbers in the interval [0, 1]. In Sugeno's model, the conclusion of the rule for output is calculated as a linear function of the inputs. Gaussian membership functions enable and ensures class transition smoothness[30,31].

The mean square error (MSE) also called as classification error is computed as:

$$\boldsymbol{mse}=\frac{1}{\boldsymbol{N}}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{y}}^{iP}-\boldsymbol{y}^{iP}\right\|^2 \tag{9}$$

The classification error for pattern is calculated as

Journal of Theoretical and Applied Information Technology
31st July 2016. Vol.89. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$mse = \begin{cases} 0, & \text{Correct Class} \\ 1, & \text{InCorreect Class} \end{cases} \quad (10)$$

Initially FIS is trained with 50 continuous gestures from Indian sign language with one sample per gesture. The training of FIS is done using least squares (LSQ) algorithm [32,33,34]. LSQ algorithm simply calculates the square error between actual outputs and targeted outputs to update the weight vector. The procedure for FIS training is shown in Figure 3.

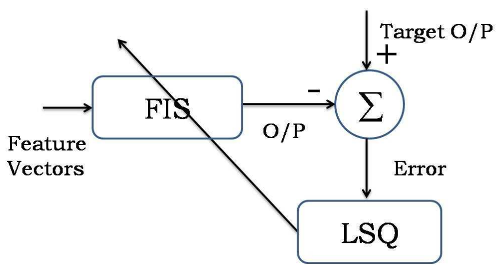

Figure 3.FIS Model used for Training

5. RESULTS AND DISCUSSION

Hand shape features are extracted in each frame with active contour level set model is characterized around 70% for any sign language. Hand shapes from each frame are represented with hand outliners extracted with shape numbers. A feature matrix is extracted between one SOS and EOS. This feature matrix becomes input to the classifier. Training to the Artificial Neural Network classifier is provided with this feature matrix. Error back propagation algorithm is used as a training algorithm. The error is calculated with gradient descent algorithm. The following methods ANN and FIS model is used for combined feature vector classification. The details of weights, biases and learning rate values are randomly initialized and re-iteratively updated with a least squares optimization of error.

For testing the above process on the images and videos of discrete and continuous sign language, a camera setup is constructed. To ensure uniform lighting, two 23 luminance bulbs are erected at an angle of 45 degrees from the signer. A HD Sony camcorder at a distance of 22 meters ensures excellent HD videos of continuous sign language. For experimentation for this model a subset of Indian Sign Language is used.

A total of 10 test subjects were used. Each performing the same set of sentences for Indian Sign Language. A discrete and continuous sentence comprising 86 words is chosen for our experimentation. The sentence used is ***"a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, Hai, Good Morning, My Name is Kishore, I am Student of K.L.University, Studying final year Undergraduate Engineering, From Department of Electronics and Communications engineering, the college is located at a lush green surroundings, with estimated an area of around 50 acres, we are doing project on sign language, thank you"***. Each video recording of the above sequence of sentences is averaged around 100 sec. This is because; the signers are not regular users of sign language.

Sony camcorder outputs videos with a 30fps HD sequences with a frame size of. For a 100 sec sequence we are looking at 3000 frames for a sentence of 86 words. For 10 different signers the figure is 30000 frames. Different signers are selected to make the system signer independent of these 5 samples are used for training and 5 will be testing samples. From any 5 samples feature vector is built combining shape outliner's vectors from AC models, which will train artificial neural network.

Hence shape information of hands should accompany the tracks to the input of the classifier. Hand shape extraction is accomplished with active contour model. The contour is placed in an area close to the head of the signer. This enables the AC segmentation algorithm to trigger only when there is significant movement near the torso of the signer. The AC segmentation process on a video frame of the signer. The contour is placed close to the torso of the signer to keep the number of iterations for segmentation to a minimum.

The segmentation is near perfect and no further processing is required. The head and hand shape boundary numbers are considered as feature vector per frame. A set of frames are presented in the sentence described above. Total numbers for frames in this particular video are around 3000. Only 699 frames have useful information that can be considered as feature vector design. These 699 frames are selected based on frame differencing model. Here threshold is set based on the velocity difference value. Large velocity value changes in frames are retained and those with lesser velocity gradient are discarded.

The contours are extracted from the region boundaries of the segments. Hand and head segment contours are manually labeled as Head and hand contours. The extracted contours for few frames are shown in fig. The boundaries of the contours are given unique combination of numbers to uniquely identify a particular shape in the video frame. These labeled shape numbers are fused with tracking features for the same frame. A final feature matrix is an amalgamation of two important characteristics for machine understanding of sign language from video sequences.

Journal of Theoretical and Applied Information Technology
31st July 2016. Vol.89. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The feature vector for both training and testing is built based on velocity vector gradients. Lesser gradients marks Star of sign (SOS) and End of Sign (EOS) frames. All the middle frames will have a feature vector. In this work we have 86 words and hence we have 86 feature vectors. Each feature vector is represented by variable number of samples with shape numbers. For a complete 86 word sentence we have 699 frame video.

The target matrix consists of 86 words in the sentence in the order of sequence described previously. To improve the efficiency of the training program, three more samples are added to the one derived feature matrix earlier. The training input vector is 200×699 whereas the target is 86×699.

Both the matrices are supplied as input to the neural network and Sugeno fuzzy inference engine with 200 inputs and 86 targets. Gradient descent error is transmitted to update the weights after every iteration as described in section 4. Log sigm oid activation function is used in all the layers. The model of neural network object is created in MATLAB. And Least squares fit error is transmitted to update the weights after every iteration as described in section 5. Fuzzy rule set is generated with feature vectors.

The performance of the proposed system is computed with word matching score given by

$$\omega_{ms} = \frac{Word\ Matching}{Total\ Words} \times 100$$

For individual words in the sentence, word matching score is computed with 5 samples for training and remaining 5 samples and the 5 already trained ones are used for testing the trained network. Table I gives values of WMS for pictures, distinct and continuous Indian sign languages videos by using ANN and FIS. The WMS values in the table-I are averaged values over 5 times. Each time training is accomplished with same training set for all models of sign language systems in table -I.

Table I. Word Matching Scores For Pictures, Distinct And Continuous Indian Sign Language Videos.

sign	pictures		Discrete videos		Continuous videos	
	Recognition Rate (%) Using active contour segmentation +ANN	Recognition Rate (%) Using active contour segmentation +FIS	Recognition Rate (%) Using active contour segmentation +ANN	Recognition Rate (%) Using active contour segmentation +FIS	Recognition Rate (%) Using active contour segmentation +ANN	Recognition Rate (%) Using active contour segmentation +FIS
a	100	100	90	90	90	90
b	90	90	90	100	90	90
c	100	90	80	90	100	90
d	100	100	90	80	90	100
e	100	90	100	90	90	90
f	90	90	90	100	90	80
g	100	90	90	100	100	60
h	90	90	100	90	100	80
i	100	100	80	90	80	90
j	90	90	90	100	100	80
k	100	100	100	90	100	100
l	90	90	90	90	90	90
m	100	90	80	100	100	80
n	100	90	90	90	90	80
o	90	80	100	90	90	100

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

p	90	100	90	90	90	100
q	90	90	100	100	100	100
r	100	100	100	90	90	80
s	90	100	90	100	90	80
t	100	90	100	90	90	90
u	90	90	90	90	100	60
v	100	90	90	90	100	80
w	90	100	90	90	90	70
x	100	100	90	100	100	100
y	90	90	90	90	90	90
z	100	90	100	100	100	80
1	100	100	100	90	90	100
2	90	90	90	90	90	90
3	90	100	90	100	100	70
4	90	100	90	90	90	80
5	90	90	90	100	100	90
6	90	90	90	90	90	70
7	90	100	90	90	90	90
8	100	90	80	100	80	90
9	100	90	100	100	90	100
10	90	90	90	90	80	90
hai	90	90	90	90	90	80
good	100	90	90	100	100	90
morning	100	90	80	100	80	90
my	100	90	90	90	100	60
name	90	90	90	90	70	80
is	90	90	90	90	80	90
kishore	100	100	90	100	100	100
I	100	100	100	100	100	100
am	90	90	90	80	90	90
a	100	90	80	90	90	80
student	90	90	90	90	100	90
of	90	90	90	90	80	90
K	100	100	90	90	100	70
L	100	90	90	90	90	90
University	90	100	100	80	100	80
Studying	90	90	80	100	70	90
final	90	90	90	90	80	100
year	90	90	90	80	80	100
Undergraduate	90	90	90	90	90	90
Engineering	90	90	90	80	80	100

Journal of Theoretical and Applied Information Technology
31st July 2016. Vol.89. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

From	100	100	90	80	70	80
Department	100	100	100	100	100	70
of	100	90	90	70	90	90
Electronics	90	90	100	90	90	90
and	90	90	90	100	70	70
Communications	100	100	90	90	80	90
engineering	80	90	90	100	90	100
the	80	90	90	90	90	80
college	100	100	90	80	90	100
is	90	90	100	80	80	90
located	90	90	100	90	80	90
at	80	90	90	90	80	100
a	100	100	90	70	90	100
lush	90	90	100	100	80	90
green	100	90	100	80	90	90
surroundings	90	100	90	90	90	100
with	100	100	100	70	80	100
estimated	100	100	90	80	100	90
area	100	90	80	90	90	80
of	80	90	80	80	80	90
around	90	90	100	80	80	100
Fifty	100	90	80	80	90	100
acres,	90	80	90	90	90	80
we	90	90	90	90	80	90
are	100	100	100	100	90	80
doing	100	90	90	90	80	90
this	90	100	90	80	70	90
sign	90	100	80	80	80	90
language	100	90	90	90	90	90
thank you	90	100	100	80	90	100
Total	94.07	93.14	91.28	90.12	89.07	87.91

6. CONCLUSION

In the present work a framework for recognizing gestures of Indian sign language was illustrated for simple backgrounds. The work is an extension to existing techniques of image and video processing models along with soft computing techniques and adapted them according to the requirements of recognition framework. In order to tackle the issue of simulation speed without losing much on recognition rates, Fuzzy Inference Systems (FIS) classifier is compared with ANN as a classifier. FIS trained with less number of samples and is around 40% faster compared to ANN. The average recognition rate obtained for FIS classifier was 90.4% which is comparatively less than that of ANN at 91.5%. With these contributions, the recognition frameworks described in this work has the potential to form the foundation for large-scale recognition for real time sign language modelling.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REFERENCES:

[1] Gaolin Fang and Wen Gao, "Large Vocabulary Contineous Sign Language Recognition Based on Trnsition-Movement Models", IEEE Transaction on Systems, MAN, and Cybernetics, 2007, Vol.37, No.1, pp 1-9.

[2] T.Starner and A.Pentland "Real-Time American Sign Language Recognition from video using Hidden Markov Models", Technical Report, MIT Media laboratory Perceptual computing section, Technical Report no.375,1995.

[3] Kishore, P.V.V.; Prasad, M.V.D.; Prasad, C.R.; Rahul, R., "4-Camera model for sign language recognition using elliptical fourier descriptors and ANN," 2015 International Conference in Signal Processing and Communication Engineering Systems (SPACES), Jan. 2015, pp.34-38.

[4] Wang, Hanjie, Xiujuan Chai, and Xilin Chen. "Sparse Observation (SO) Alignment for Sign Language Recognition." *Neurocomputing* 175 (2016), pp. 674-685.

[5] Tripathi, Kumud, and Neha Baranwal GC Nandi. "Continuous Indian Sign Language Gesture Recognition and Sentence Formation." *Procedia Computer Science* 54 (2015), pp.523-531.

[6] Gulhane, Vaishali M., Madhavi S. Joshi, and Manik D. Ingole. "Neural Network based Hand Gesture Recognition." *International Journal of Advanced Research in Computer Science* 3.3 (2012).

[7] Ahmad Zaki Shukor, "A New data glove approach for malaysian sign language detection", 2015 IEEE International Symposium on Robotics and Intelligent Sensors, 2015, pp.60-67.

[8] Liu Zhigui, Junbo Wang, and Yuyu Zhu. "A Study of Active Contour Segmentation Models based on Automatic Initial Contour." *International Journal of Signal Processing, Image Processing and Pattern Recognition* 8.4 (2015), pp.201-214.

[9] G. Fang, W. Gao, X. Chen, C. Wang, and J. Ma, "Signer-Independent Continuous Sign Language Recognition Based on SRN/HMM," Proc. Gesture Workshop, 2001, pp. 76-85.

[10] Y. Nam and K. Y. Wohn. Recognition of space-time hand-gestures using hidden Markov model. ACM Symposium on Virtual Reality Software and Technology, 1996.

[11] P. Viola and M. Jones, "Robust Real-Time Object Detection," Proc. IEEE Workshop Statistical and Computational Theories of Vision, 2001.

[12] L. Wu, S.L. Oviatt, and P.R. Cohen, "Multimodal Integration—A Statistical View," IEEE Trans. Multimedia, 1999, vol. 1, no. 4, pp. 334-341.

[13] E.-J. Holden and R. Owens, "Visual Sign Language Recognition," Proc. Int'l Workshop Theoretical Foundations of Computer Vision, 2000, pp. 270-287.

[14] D. H. Anderson and L. 0. Hall, "MR. FIS: Mamdani Rule Style Fuzzy Inference System"

[15] M.-C. Su, "A Fuzzy Rule-Based Approach to Spatio-Temporal Hand Gesture Recognition," IEEE Trans. Systems, Man, and Cybernetics, Part C: Application Rev, May 2000, vol. 30, no. 2, pp. 276-281.

[16] Ginmo Chung and Luminita Vese. (2003) *"Image segmentation using a multilayer level set approach.Technical Report*", UCLA, pp.03-53.

[17] R. Malladi, J.A. Sethian, and B.C. Vemuri. (1995) *"Shape modeling with front propagation: A level set approach*". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.17, No. 2, pp.158-175.

[18] D. Terzopoulos and K. Fleischer (1988) "*Deformable models"*. The Visual Computer, Vol.4 No.6, pp.306-331.

[19] D. Terzopoulos, J. Platt, A. Barr, and K. Fleischer. (1987) "*Elastically deformable models*" In Comp.Graphics Proc., ACM Press/ ACM SIGGRAPH, pp.205-214.

[20] 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 pp.271-293.

[21] Chan.T, Vese.L.A, Active contours without edges. IEEE Trans. Image Process. 2001, Vol.10, no.2, pp.266–277.

[22] D. Mumford and J. Shah. (1989) *"Optimal approximation by piecewise smooth functions and associated variational problems"* Comm. Pure Appl. Math, Vol.42, pp.577-685.

[23] M.Kass, A Witkin, D Terzopoulos, "Snakes:active Contour Models", *Int. J. of Computer Vision,* 1987,pp 321-331.

[24] Osher, S., Sethian, J. (1988) Fronts propagating with curvature-dependant speed: algorithms based on Hammilton-Jacobi formulations.J. Comput. Phys. Vol.79 No.1, pp.12–49.

[25] L.A. Zadeh, "Fuzzy Sets", Information and Control, 1965.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[26] Michael, N., 2005 Artificial Intelligence: A Guide to Intelligent Systems, Pearson Education Limited.
[27] M. Sugeno, "An introductory survey of fuzzy control," Inform. Sci., 1985, vol. 36, pp. 59-83.
[28] E. Cox, "Adaptive Fuzzy Systems," IEEE Spectrum, Feb. 1993, pp. 27-31..
[29] L. X. Wang and I. M. Mendel, "Generating fuzzy rules by learning from examples," IEEE Trans. Syst., Man, Cyben., 1992, vol. 22, no. 6, pp. 1414-1427.
[30] 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, 1920,Feburary,2011, IEEE Catalog Number: CFP1126M-PRT,pp.105-110.
[31] 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, vol.5 ,pp.479-503.
[32] Levenberg, K., "A method for the solution of certain nonlinear problems in least-squares," Quart. Appl. Math. II, 1944, pp.164–168.
[33] Marquardt, D. W., "An algorithm for least-squares estimation of nonlinear parameters," SIAM J. Appl. Math., 1963, Vol. 11, pp.431–441.
[34] Dennis, J. E., State of the Art in Numerical Analysis, Academic Press, 1977.