A Novel Approach To American Sign Language Recognition Using MAdaline Neural Network

Sriparna Saha[1], Rimita Lahiri[2], Amit Konar[3]
[1,2,3]Electronics & Tele-Communication Engineering Department
[1,2,3]Jadavpur University, Kolkata, India
[1]sahasriparna@gmail.com, [2]rimita.lahiri@yahoo.com, [3]konaramit@yahoo.co.in

Atulya K. Nagar
Mathematics and Computer Science Dept.
Liverpool Hope University
United Kingdom
nagara@hope.ac.uk

***Abstract*—Sign language interpretation is gaining a lot of research attention because of its social contributions which is proved to be extremely beneficial for the people suffering from hearing or speaking disabilities. This paper proposes a novel image processing sign language detection framework that employs MAdaline network for classification purpose. This paper mainly highlights two novel aspects, firstly it introduces an advanced feature set comprising of seven distinct features that has not been used widely for sign language interpretation purpose, more over utilization of such features negates the cumbersome step of cropping of irrelevant background image, thus reducing system complexity. Secondly it suggests a possible solution of the concerned problem can be obtained using an extension of the traditional Adaline network, formally termed as MAdaline Network. Although the concept of MAdaline network has originated much earlier, the provision of application of this framework in this domain definitely help in designing an improved sign language interpreting interface. The newly formulated framework has been implemented to recognize standardized American sign language containing 26 English alphabets from 'A' to 'Z'. The performance of the proposed algorithm has also been compared with the standardized algorithms, and in each case the former one outperformed its contender algorithms by a large margin establishing the efficiency of the same.**

Keywords—American sign language; MAdaline; Adaline; Widrow-hoff's learning rule; least mean square

I. Introduction

The primary motif behind the concept of Man Machine Interaction (MMI) is to develop an efficient interactive platform that can bridge the gap between an artificially designed machine environment and human beings [1]. Researchers have proposed numerous possible modalities to serve the purpose and hand gesture is a very popular communicative interface that is employed for this research. Hand gesture recognition is gaining a lot of research attention these days because of its diverse scopes of applications including remote controllers, media players, robot, sign language discrimination and so on. Hand gesture recognition in this context typically refers to recognizing the non-verbal body actions involving the movement of hand, face, shoulder, arm etc. Hand gesture recognition can be broadly categorized into two distinct approaches, firstly the vision based approach employs camera like devices to capture human motion and associated information, but occurrence of background and undesirable obstructions require efficient processing techniques for further analysis. Contrarily, glove based systems rely upon the sensors attached to the gloves to record the hand movements and additional data required [2].

With the widespread use of computers sign language interpretation has been emerged as one of the most promising research sub-disciplines in this vast domain of hand gesture recognition as the limitations of the traditionally used interfaces involving mouse and keyboard became more and more pronounced. A sign language typically refers to a way of interaction by visually transmitting the sign patterns embodied by a subject to convey an intended message. It is an efficient communication tool for speech and hearing impaired people, so the main motivation for developing such a system lies in the contribution of the system towards increasing social awareness and worth of it in providing social aid. The main objective of sign recognition is to build a robust and reliable framework of transcribing sign languages into text or speech which in turn facilitates efficient communication between deaf and dumb people. Sign language recognition is an ill-posed problem because sign languages are multichannel, conveying different meanings at once, moreover, they are highly inflected that is the appearance of the same sign is likely to take different appearances depending on the subject enacting it.

There are numerous sign languages available across the globe as sign languages are greatly influenced by local culture and native languages specific to that place. All these sign languages differ greatly in terms of syntax, morphology, phonology and grammar as well. In the present work, only American Sign Language (ASL) has been addressed.

Literature shows a lot of research work has already been carried out in the field of sign language recognition. Vogler *et al.* [3] designed an HMM based American Sign Language recognition system that uses a context dependent model to deal with the issue of co-articulation that is to deal with simultaneous changes like involvement of both hands or hand shape changes. The authors reported that use of three dimensional features have surpassed the performance yielded using two dimensional features proving the worth of the former one. Wilson *et al.* [4] proposed a novel method of parameterized gesture recognition by extending the concepts of classical HMM after incorporating a global parametric variation in the output states of HMM leading to an expectation maximization algorithm based training and testing scheme and the authors also inferred about the superiority of Parametric HMM (PHMM) over HMM in terms of recognition rate. Brashear *et al.* [5] suggested the use of HMM to develop an ASL based game CopyCat and here it is important to note that color histogram based adaptation has been employed to build robust hand segmentation and partition, moreover accelerometer data along with hand shape information has been fed to HMM for data analysis. Chen *et al.* [6] designed a novel gesture recognition paradigm by partitioning the processing steps into four sections, a real time hand tracking tool to trace the moving hand and extract it from the background, a Fourier Descriptor and Motion Analysis to characterize spatial and temporal features respectively, finally a merged feature vector comprising of specimen of spatial and temporal features is fed to an HMM for training purpose and later trained HMM is utilized to recognize the gestures.

Ren *et al.* [7] emphasized on the need of designing a robust Kinect based hand gesture recognition interface by introducing a novel distance metric for hand dissimilarity measure termed as "Finger Earth Mover's Distance (FEMD)" to address the issue of obtaining noisy hand shape from Kinect sensor. Chen *et al.* [8] delved further into the depth of the topic and came up with an innovative scheme of real time gesture recognition following a two level approach. The lower level of the approach implements the posture recognition with Haar-like features and the AdaBoost learning algorithm while higher level implements the linguistic hand gesture recognition using a context-free grammar-based syntactic analysis. Chuan *et al.* [9] introduced an economical and portable alternative of Microsoft Kinect sensor known as palm sized Leap motion sensor for designing a real time gesture interpreting system and the researchers have employed the classical k-nearest neighbor (k-NN) and support vector machine (SVM) algorithm to conduct the classification phase with a fairly good recognition rate.

As discussed above, although researchers have suggested numerous approaches for dealing with the issues faced for sign language interpretation, none of them provided optimal solutions satisfying all constraints. This motivated us to revisit the classical concepts associated with pattern recognition based systems and come up with an improvised idea with concrete theoretical background that provides solution to this problem. This paper emphasizes on the use of neural network classifiers assuming the gestures to comprise of non-linear components. In the present work, MAdaline network has been employed to take care of the recognition phase. More importantly, this work opens a new avenue of obtaining sophisticated feature descriptors depending upon multiple raw features. Obviously, these new set of features are likely to provide the most relevant information from all aspects as compared to other features. The mentioned framework has been implemented for detection of sign language corresponding to the 26 distinct alphabets of English literature.

The rest of the paper is divided into four sections, section II recapitulates the postulates of MAdaline network with necessary diagrams, section III presents a detailed outline of the proposed system, section IV presents the experimental results after describing the set up and finally section V draws the necessary conclusion of the present work.

II. MAdaline Neural Network

The basic building block employed in most of the neural networks is an Adaptive Linear Element popularly known as Adaline [10]–[12] network. Mathematically, an Adaline neural network can be defined as an early stage artificial neural network that accomplishes classification by altering the network weights at each iteration with an aim to minimize the mean square error to the maximum possible extent. Currently, Adaline networks find extensive use in the field of pattern recognition as well as adaptive filtering. Basically, it is the result of cascading a adaptive linear combiner and a hard limit quantizer producing a binary output. Similar to the perceptron, Adaline also employs a threshold logic device that performs a linear summation of the inputs as given in Fig. 1. The main highlight of this network is the implementation of Widrow-hoff's learning rule in the basic adaptive linear neuron model.

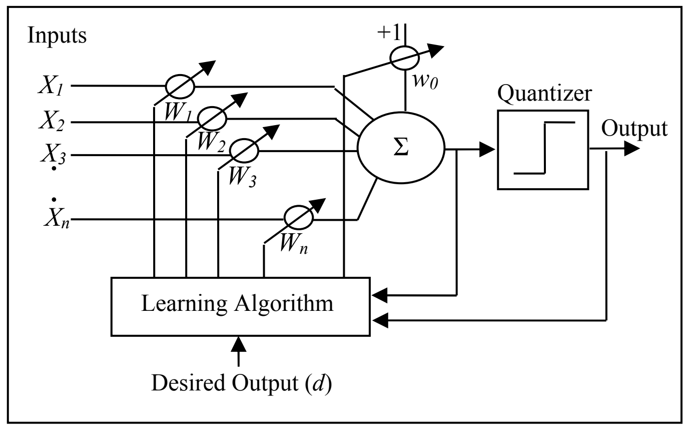

Fig. 1. Architecture of Adaline neural network.

The Widrow-Hoff's learning rule [13] computes the value of the error function at each iteration and adapt the weights accordingly to eliminate the error as far as possible in Least Mean Square (LMS) sense. The Widrow–Hoff rule for adapting the weight of the *i*-th neuron is expressed as,

$$\Delta W_i(t+1) = \eta[d(t) - \sum_{i=1}^{n} W_i(t) X_i(t)] X_i(t) \tag{1}$$

where η denotes the learning rate such that $0 \le \eta \le 1$ and $\eta \le \frac{1}{n}$, $W_i(t)$ and $X_i(t)$ denote the *i*-th weight and *i*-th input pattern at interaction *t*. Similarly, $d(t)$ denotes the target output at iteration *t*. Finally, the squared error can defined as,

$$E = [d(t) - \sum_{i=1}^{n} W_i(t) X_i(t)]^2 \tag{2}$$

For optimal results, the weights are required to be altered in the direction of negative gradient that is $-\frac{\partial E}{\partial W_i}$. It is evident from (2), that decomposing it E can be expressed as,

$$E = d^2(t) - 2d(t)\sum_{i=1}^{n} W_i(t)X_i(t) + [\sum_{i=1}^{n} W_i(t)X_i(t)]^2 \quad (3)$$

Now, calculating the gradient E with respect to W_i yields,

$$\frac{\partial E}{\partial W_i} = -2[d(t) - \sum_{i=1}^{n} W_i(t)X_i(t)]X_i(t) \quad (4)$$

So, the weights should be altered by an amount (1), and finally, the adapted weights can be expressed as,

$$W_i(t+1) = W_i(t) + \Delta W_i(t+1) \quad (5)$$

$$W_i(t+1) = W_i(t) + \eta[d(t) - \sum_{i=1}^{n} W_i(t)X_i(t)]X_i(t) \quad (6)$$

Typically, the usage of Widrow–Hoff learning rules assure minimization of local error and experimentally it is determined that a typical Adaline model is expected to converge in five times as many number of trials as there are learning weights.

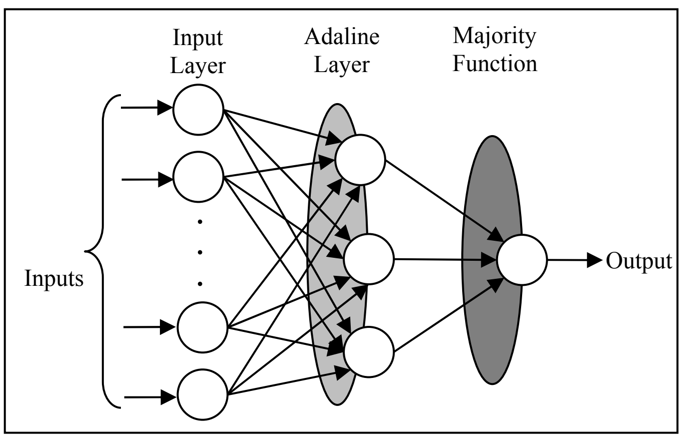

Fig. 2. Architecture of MAdaline neural network.

Despite its efficiency in reducing the error value, single Adaline networks are not capable of executing non-linear classification task. To enable non-linear classification the single Adaline network based systems are extended further to include multiple Adaline neurons organized in multi-layer structure as shown in Fig. 2. The basic working principle of each Adaline unit remains unchanged and final output is derived by applying majority voting technique upon the outputs of Adaline layer. Since the present work deals with gestures where there is every possible chance of occurrence of non-linearity, so Multiple Adaline (MAdaline) framework has been chosen over classical Adaline neurons. Table I presents the pseudocode for the application of MAdaline network in the present problem context, here M is the number of Adaline units employed, K is the number of classes and n is the dimension of the input.

TABLE I. ALGORITHM FOR MADALINE NETWORK

```
Encode the K target d_K outputs  in the range [-1,+1]
Training:
For iteration= 1to Max
        For  i=1to M
                For j=1to n
                        Randomly initialize the weights
                End
        End
        For i=1 to M
                Present the input weights X_i1, X_i2,.... X_in
                Calculate the Actual Adaline Outputs for all the K
                output classes,Y_i1,Y_i2,....,Y_iK.
        End
        Calculate the actual MAdaline Output for each of the K
        classes as, S_1, S_2,....S_K.
        S=Majority voting(Y_1, Y_2,....Y_M)
        Determine the error value and adapt the weight of the
        winning neuron using delta rule.
End
Testing:
Present the unknown sequence to the network and obtain the K outputs
and depending upon the closeness of the actual output to the target
outputs, class label is obtained.
```

III. OUTLINE OF THE PROPOSED WORK

The details of the proposed work are explained here.

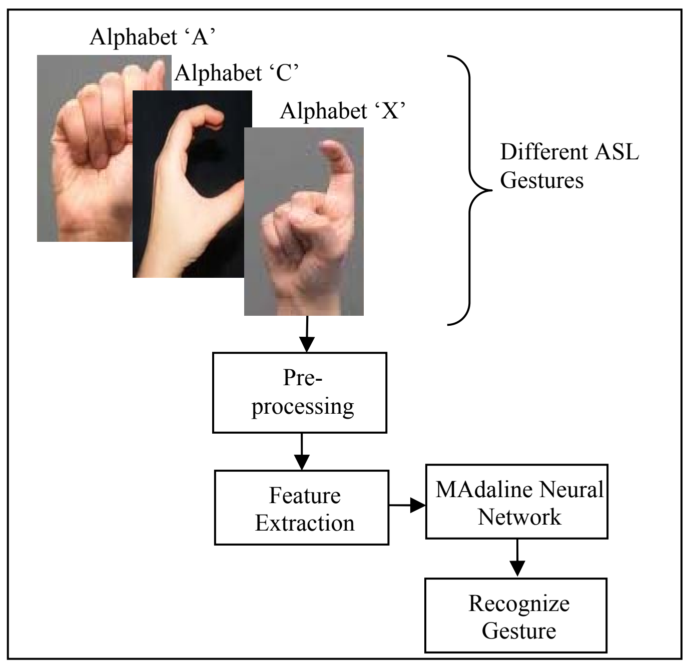

Fig. 3. Block diagram of the proposed work.

A. Pre-processing Stage

The significant part of hand gesture recognition is to identify the hand from the background. The actual input image is a part of RGB color space, hence invariably suffers from the drawbacks of high correlation, non-uniformity and mixing of chrominance and luminance. To outdo these issues RGB images are transformed into HSV and YCbCr color spaces as skin color is comprised of a certain combination of *H*, *Cb* and *Cr* values.

$$Cb = 0.15R - 0.29G + 0.44B + 128 \tag{7}$$

$$Cr = 0.44R - 0.37G - 0.07B + 128 \tag{8}$$

$$H = \arctan\left(\frac{\sqrt{3}\left(G-B\right)}{\left(R-G\right)+\left(R-B\right)}\right) \tag{9}$$

The thresholds are determined experimentally $140 \leq Cb \leq 195$, $140 \leq Cr \leq 165$ and $0.01 \leq H \leq 0.1$ relying upon the skin color values. Once the segmentation is done, the image is converted into binary space to execute morphological dilation. Further, the inner holes of the images are filled to have a smoothening impact. The entire process is depicted in Fig. 4.

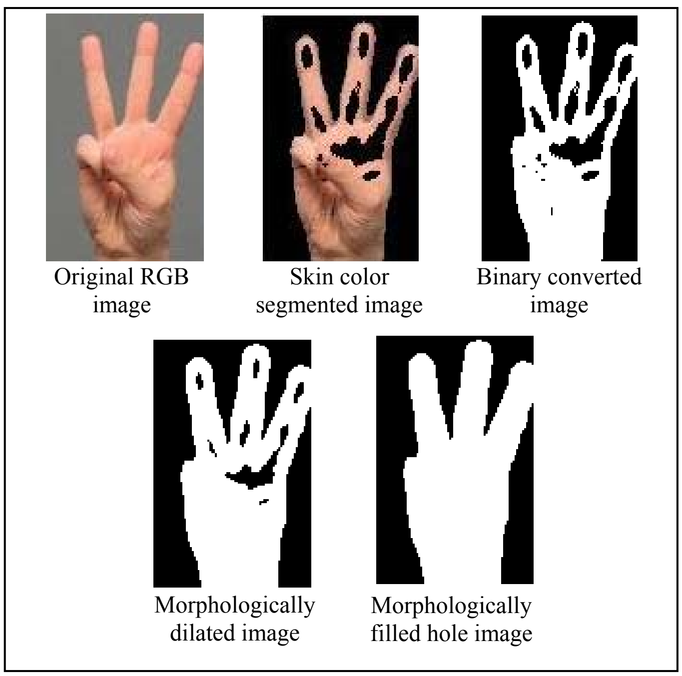

Fig. 4. Results obtained through different stages of pre-processing.

B. Feature Extraction Stage

For this proposed work, we first extract four raw features from each pre-processed hand gesture. Then based on these raw features, we have calculated seven final sophisticated features [14] required to perform gesture recognition using MAdaline neural network. The reason behind deriving these two types of features is that the distance of the hand gesture from the camera can vary widely. Also we have developed this work for non-cropped images of hand gestures. Thus we need to extract features which is not going to vary based on the above two criteria.

The first two raw features are largest length (*L*) and largest width (*W*) possible within the binary gesture. The illustrations of these two features are shown in Fig. 5. The third feature is the amount of area (*A*) embedded within the hand gesture. This feature is measured by summing the total number of 1 binary value present in the pre-processed gesture. Last raw feature is the perimeter (*P*) by accounting the total number of pixels are present in the gesture boundary.

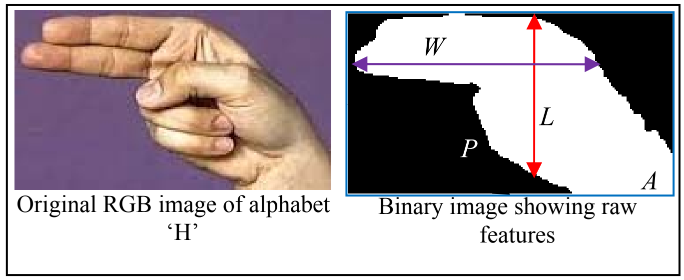

Fig. 5. Depiction of raw features obtained from pre-processed gesture.

From the combination of these 4 raw features, we have measured 7 features, which are given in Fig. 6.

a) Aspect ratio: The first feature is the aspect ratio which is the ratio between *L* and *W*.

b) Circularity: This describes the similarity between a circle and the gesture. It is calculated by the ratio between $4\pi A$ and P^2.

c) Rectangularity: This feature is for measuring the closeness between a rectangle and the getsure. It is measured by the ration between ($L \times W$) and *A*.

d) Ratio of perimeter to length and width: It is the ratio between *P* and (L+W).

e) Ratio of perimeter to area: It is the ratio between *P* and *A*.

f) Ratio of length to area: It is the ratio between *L* and *A*.

g) Ratio of width to area: It is the ratio between *W* and *A*.

L/W	$4\pi A/P^2$	LW/A	$P/(L+W)$	P/A	L/A	W/A

Fig. 6. Seven features obtained for training of MAdaline neural network.

C. Application of MAdaline Neural Network for ASL Recognition

For executing the present work, here 10 Adaline units are considered (M=10), as mentioned earlier since the framework is tested for 26 different alphabets, so K=26. Since seven distinct features are extracted for this work, we assume the 7 input patterns to each Adaline unit (n=7). While running the source codes maximum iteration is taken to be 2000 (*Max*=2000).

IV. Experimental Results

For the preparation of training dataset, we have accumulated 30 instances for each 26 hand gestures of ASL. Thus total images present in the training dataset are 780 (=26×30). Again for the testing purpose, altogether different 200 images that do not belong to the training dataset are taken into account. The sizes of all the images belong to both training and testing phases are not of equal size and also the images are not cropped. The calculation procedure for features is given in Table II. Firstly, the raw features as stated in Table II are determined and in the next step the final feature set is obtained where each feature is calculated following a mathematical relation between more than one raw feature. It is important to note that the improvised feature set is more advanced as using these features , it is not necessary to use cropped image to deemphasize the unimportant background of the image.

The performance of the proposed method using Madaline neural network is compared with Finger-Earth mover's distance (FEMD), k-nearest neighbor (kNN) and support vector machine (SVM), back propagation neural network (BPNN), recurrent neural network (RNN), probabilistic neural network (PNN) and feed forward neural network (FFNN). As we have used static gestures, thus we are not comparing it with HMM. The performance metrics are given in (10-14).

$$\text{Precision} = \frac{TP}{TP+FP} \tag{10}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{11}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{12}$$

$$F1\ Score = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{13}$$

Here *TP*, *FN*, *FP* and *TN* denote True Positive, False Negative, False Positive and True Negative. The results achieved from performance analysis are given in Fig. 7 to 9. The computational complexity values achieved for the above mentioned algorithms are in second unit for a Windows 7 PC with 2GB RAM. Based on all the performance analysis, it is evident that MAadaline neural network is the best choice for recognition of ASL hand gestures.

To statistically justify our results, we have used McNemar's statistical test [15]. Let we have two classifiers *A* and *B*, then n_{01} be the number of examples misclassified by f_A but not by f_B, and n_{10} be the number of examples misclassified by f_B but not by f_A.

$$Z = \frac{\left(\left|n_{01}-n_{10}\right|-1\right)^2}{n_{01}+n_{10}} \tag{15}$$

In Table III, the null hypothesis has been rejected, if Z> 3.84, where 3.84 is the critical value of χ2 for 1 degree of freedom at probability of 0.05.

TABLE II. Features Extraction from Unknown Gesture

Unknown RGB image	Raw features		Required features	
	L	212	L/W	1.2544
	W	169	$4\pi A/P^2$	0.4366
	A	1.5182e+04	LW/A	2.3599
	P	661	$P/(L+W)$	1.7349
			P/A	0.0435
			L/A	0.0140
			W/A	0.0111

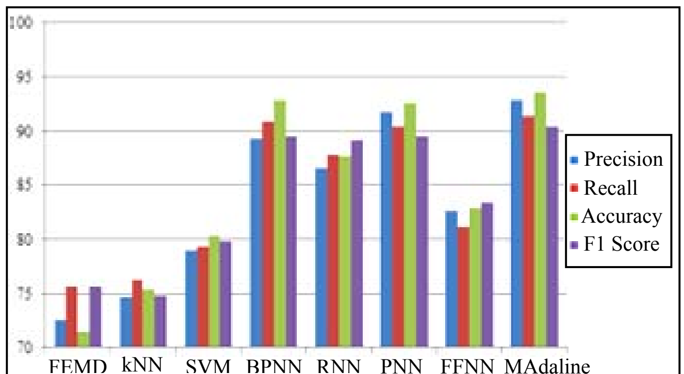

Fig. 7. Comparison of proposed work with existing literatures.

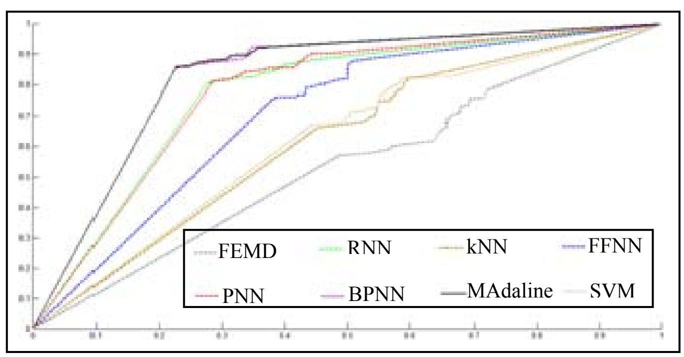

Fig. 8. ROC curves for the competitive algorithms.

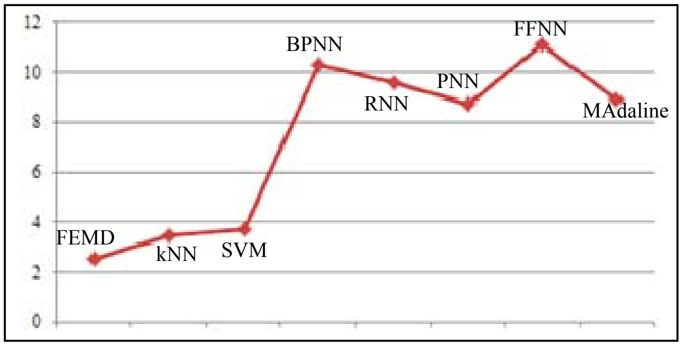

Fig. 9. Computational complexity values for the competitive algorithms.

TABLE III. MCNEMAR'S STATISTICAL TEST

	f_A=MAdaline			
f_B	n_{01}	n_{10}	Z	Comment
FEMD	8	41	20.8980	Reject
kNN	6	37	20.9302	Reject
SVM	14	29	4.5581	Reject
BPNN	21	25	0.1957	Accept
RNN	16	41	10.1053	Reject
PNN	17	34	5.0196	Reject
FFNN	12	44	17.1607	Reject

V. CONCLUSION

This paper presents an innovative configuration of sign language recognition from the image processing perspective. Four findings obtained for this present paper. Firstly, the proposed algorithm does not require an additional step for extracting the relevant image information from the background. Secondly, although the features derived here require elementary geometrical concepts, but these features are more sophisticated and easy to deal with than the popularly used geometric features. Thirdly, computation of such features involves processing of multiple raw features and as an obvious consequence these features are richer in terms of information content. Finally, high classification accuracy obtained using MAdaline network clearly justifies the usage of such networks in this domain. This work introduces an innovative set of feature descriptors for this problem, but the susceptibility of these feature set to carry redundant information should also be checked to avoid computational overhead, further this framework should also be tested for other sign languages like Indian sign language, Chinese sign language and so on, and combining all these a locality independent sign language recognition system can be designed.

ACKNOWLEDGMENT

The research work is supported by the University Grants Commission, India, University with Potential for Excellence Program (Phase II) in Cognitive Science, Jadavpur University and University Grants Commission (UGC) for providing fellowship to the first author.

REFERENCES

[1] J. Preece, Y. Rogers, H. Sharp, D. Benyon, S. Holland, and T. Carey, *Human-computer interaction*. Addison-Wesley Longman Ltd., 1994.

[2] R.-H. Liang and M. Ouhyoung, "A real-time continuous gesture recognition system for sign language," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, 1998, pp. 558–567.

[3] C. Vogler and D. Metaxas, "Handshapes and movements: Multiple-channel american sign language recognition," in *Gesture workshop*, 2003, vol. 2915, pp. 247–258.

[4] A. D. Wilson and A. F. Bobick, "Parametric hidden markov models for gesture recognition," *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 21, no. 9, pp. 884–900, 1999.

[5] H. Brashear, V. Henderson, K.-H. Park, H. Hamilton, S. Lee, and T. Starner, "American sign language recognition in game development for deaf children," in *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, 2006, pp. 79–86.

[6] F.-S. Chen, C.-M. Fu, and C.-L. Huang, "Hand gesture recognition using a real-time tracking method and hidden Markov models," *Image Vis. Comput.*, vol. 21, no. 8, pp. 745–758, 2003.

[7] Z. Ren, J. Yuan, and Z. Zhang, "Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera," in *Proceedings of the 19th ACM international conference on Multimedia*, 2011, pp. 1093–1096.

[8] Q. Chen, N. D. Georganas, and E. M. Petriu, "Real-time vision-based hand gesture recognition using haar-like features," in *Instrumentation and Measurement Technology Conference Proceedings, 2007. IMTC 2007. IEEE*, 2007, pp. 1–6.

[9] C.-H. Chuan, E. Regina, and C. Guardino, "American Sign Language recognition using leap motion sensor," in *Machine Learning and Applications (ICMLA), 2014 13th International Conference on*, 2014, pp. 541–544.

[10] B. Widrow and M. A. Lehr, "30 years of adaptive neural networks: perceptron, madaline, and backpropagation," *Proc. IEEE*, vol. 78, no. 9, pp. 1415–1442, 1990.

[11] R. C. Eberhart, *Neural network PC tools: a practical guide*. Academic Press, 2014.

[12] M. M. Nelson and W. T. Illingworth, *A practical guide to neural nets*, vol. 1. Addison-Wesley Reading, MA, 1991.

[13] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. De Jesús, *Neural network design*, vol. 20. PWS publishing company Boston, 1996.

[14] S. G. Wu, F. S. Bao, E. Y. Xu, Y.-X. Wang, Y.-F. Chang, and Q.-L. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network," in *Signal Processing and Information Technology, 2007 IEEE International Symposium on*, 2007, pp. 11–16.

[15] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Comput.*, vol. 10, no. 7, pp. 1895–1923, 1998.