e-ISSN : 0976-5166
p-ISSN : 2231-3850

PERFORMANCE COMPARISON OF SVM, CNN, HMM AND NEURO-FUZZY APPROACH FOR INDIAN SIGN LANGUAGE RECOGNITION

Ms. Hemina Bhavsar

Sardar Patel University,
Vallabh Vidyanagar, *388120,* India,
heminabhavsar@gmail.com

Dr. Jeegar Trivedi

Sardar Patel University,
Vallabh Vidyanagar, *388120,* India,
jeegar.trivedi@gmail.com

Abstract

Hearing Impaired or mute peoples are uses sign language to express their thoughts in front of each other as well as normal people. This research paper describes proposed methodology for Indian Sign language recognition where images of alphabets signs are used for recognition. Various image processing techniques have been applied to smooth and filter the images. Similarity index values of testing data and training data have been found as a feature using correlation-coefficient algorithm. This paper consists of a comparison of classification algorithms: Support Vector Machine (SVM), Convolutional Neural Network (CNN), Hidden Markov Model (HMM), and Neuro-Fuzzy(NF) approach. Comparison performs by performance evaluation on MATLAB. Total 200 images of alphabets A to J are tested where 100 images are positive and remaining 100 images are negative. Testing results of positive images consists of accuracy, 94% for SVM, 70% for HMM, 95% for CNN and 97% for NF. Comparison of NF with SVM, HMM, CNN is also describing for the different parameters. Performance is calculated by the confusion matrix where NF approach consists of 96% accuracy.

***Keywords*: Convolutional Neural Network (CNN), Hidden Markov Model (HMM), Indian sign language (ISL), Neuro-Fuzzy (NF), Support Vector Machine (SVM)**

1. Introduction

Deaf & dump people use sign language to communicate with each other. Because of they are unable to speak or unable to hear, they use hand gesture with facial expression to express the sign. Variety of sign languages is used in different countries like Indian sign language (ISL), Pakistani Sign Language (PSL), American Sign Language. Each type of sign languages has different type of sign from each other. In this paper we described our focus on ISL.

Normal People have faced difficulties to understand the sign language. Due to lack of human translator we need some automatic machine translator that can be able to convert sign into normal language. This kind of translator is called as sign language recognition system.

Sign language recognition is an automatic computerized system that taken the gesture formats and converts it into normal language text. In this paper we described the discussion & comparison of various classification algorithms which are used to develop the sign language recognition system. We tested algorithms on MATLAB& conclude best solution.

2. Related Study

Ashok Kumar Sahoo, Kiran Kumar Ravulakollu [1] presented on Indian Sign Language Recognition using Skin Color Detection, theiremphasis is on the training and testing datasetbuild, feature extraction methods with the use of skin color detection and classification is done by k-Nearest Neighbor and Neural Network classifiers for Indian Sign Language limited words. The skin color detection method uses to segment the hand and face port of images from

Input video frames. With this kind of framework, they obtained accuracy from k-Nearest Neighbor classifier is 97.04% with image pixel feature extraction method. An accuracy rate of 97.00% obtained in combination of image pixel feature extraction method and Neural Network classifier.

Madhuri Sharma, Ranjna Pal and Ashok Kumar Sahoo [2] presented on INDIAN SIGN LANGUAGE RECOGNITION USING NEURAL NETWORKS AND KNN CLASSIFIERS, they utilized direct pixel value and hierarchical centroid techniques to get the features from sign pictures. In the wake of extracting features from images, neural network and kNN classification algorithms were utilized to order the signs. The consequence of these analyses is accomplished up to 97.10% exactness by testing of 5000 different numeric signs.

Aliaa A. A.Youssif, Amal ElsayedAboutabl, Heba Hamdy Ali [3] presented on Arabic Sign Language (ArSL) Recognition System Using HMM. They presents a programmed Arabic sign language recognition system depended on the Hidden Markov Models (HMMs). A huge arrangement of tests has been utilized to perceive 20 secluded words from the Standard Arabic gesture based communication. The proposed framework is signer independent. Tests are led utilizing genuine recordings taken for hard of hearing individuals in various dresses and with various skin hues. Our framework accomplishes a general acknowledgment rate coming to up to 82.22%.

L. Pigou describe the utilization of CNN's to characterize 20 Italian motions. They focus on person gesture spotting rivalry [4]. They utilize a Microsoft Kinect on full body images of individuals playing out the gestures and got the 91.7% accuracy with the used of CNN. They mentioned utilization of 3-D glove, the Kinect sensor catch feature, which helps fundamentally in arranging ASL signs.

Sharma et al. utilize the individual classifiers (Support Vector Machines and k-Nearest Neighbors) to describe each color format by subtract the background color and smooth the image by removal the unwonted noise [5]. Their development originates from utilizing a contour follow, which is a productive portrayal of hand shapes. They achieve a precision of 62.3% utilizing a SVM on the sectioned shading channel model.

Admasu and Raimond recognize the Ethiopian Sign Language. They got 98.5% accuracy by utilizing a feed forward Neural Network [6]. They utilize a lot of picture preprocessing. They found features with a Gabor Filter and Principal Component Analysis.

P. V. V. Kishore , P. Rajesh Kumar [7] presented on A Video Based Indian Sign Language Recognition System (INSLR) Using Wavelet Transform and Fuzzy Logic. They proposed a wavelet based video division strategy to classify sign language based hand gesture. Shape features are extracted by utilizing circular Fourier depictions. Gestures recognition done from the extracted features iby using Sugeno type fuzzy inference system. With the use of this framework they tested 80 different words and got 96% accuracy.

GaolinFang, Wen Gao, and Debin Zhao [8] exhibited on Large Vocabulary Sign Language Recognition Based on Fuzzy Decision Trees. Two hands classifier and a hand-molded classifier with minimal computational expense are first used to dynamically take out numerous outlandish up-and-comers, and afterward, a self-sorting out element maps/shrouded Markov model (SOFM/HMM) classifier in which SOFM being as a certain various endorsers' element extractor for constant HMM, is proposed as an extraordinary segment of a fluffy choice tree to get the conclusive outcomes at the last non leaf hubs that just incorporate a couple of competitors. Test results on an enormous jargon of 5113-signs show that the proposed strategy drastically diminishes the acknowledgment time by multiple times and furthermore improves the acknowledgment rate about 0.95% over single SOFM/HMM.

Grobel and Assan [9] utilized HMM to perceive segregated signs with 91.3% precision out of a 262-sign jargon. They separated two-dimensional (2-D) highlights from video accounts of endorsers wearing hued gloves.

HMM was likewise utilized by Hienz and Bauer [10] to perceive video based German gesture-based sign language with a solitary shading camcorder as information. Their examination was an expansion of the work by Grobel and Assan. A precision of 91.7% can be accomplished in acknowledgment of gesture-based communication sentences with 97 signs.

Liang and Ouhyoung [11] utilized the time-differing parameter edge of hand stance to decide end-focuses in a flood of motion contribution for ceaseless Taiwan SLR with the normal acknowledgment pace of 80.4% for 250 signs. In their framework, a Dataglove was utilized as an info gadget, and HMM was taken as acknowledgment technique.

Starner et al. [12] utilized a view-based methodology for constant American SLR. They utilized single camera to separate 2-D highlights and the removed highlights were then taken as the contribution of HMM. The word exactness of 92% or 98% was gotten when the camera was fitted on the work area or in a client's top in perceiving the sentences with 40 distinct signs.

SabahetaĐogić ,Gunay Karli [13] presented on Sign Language Recognition using Neural Networks. The work is finished with the utilization of advanced picture preparing techniques giving a framework that shows a multilayer neural system utilizing a back propagation algorithm. Pictures are handled by feature extraction techniques, and by veiling strategy the informational index has been made. Preparing is finished utilizing cross approval strategy for better execution hence; an exactness of 84% is accomplished.

Author	Algorithm and Methods used	Description	Accuracy
[14]	CNN	ISL recognition of Numbers by Single Hand using Vision Based Approach	99.56% and in low light 97.26%
[15]	CNN, RNN, Prediction and Pool Layer Approach	Argentinean Sign language recognition of words by Single handed and Double Handed using Vision Based Approach	95.20%
[16]	Support Vector Machines	Real Time Hand Gesture recognition System for Android Devices	93%
[17]	Support Vector Machines	LIBRAS Sign Language Hand Configuration Recognition	96%
[18]	Hidden Markov Model, Dynamic Time Wrapping	Sign Language Recognition	89%
[19]	Hidden Markov Model, Dynamic Time Wrapping	Real-time Ukrainian sign language recognition system	91.70%
[20]	feedforward backpropagation of ANN	American Sign Language Recognition	95%

Table 1. Comparison of Related work

3. Design Of Proposed System Architecture

3.1. System Architecture

The architecture of proposed system is based on HMM/SVM/CNN/NF classification algorithms. As shown in following fig.1. Sign language images are taken by input device and feed into proposed architecture. First pre-processing applied on taken images using image processing, then proceed to feature extraction module and then classification is performed using HMM, SVM, CNN, and FN respectively. This architecture is implemented on Matlab.

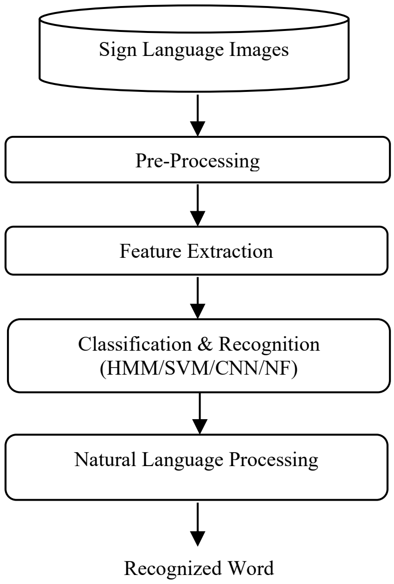

Fig 1. Architecture of Proposed System

3.2. Data Collection

We trained a database with self generated images. Trained database is a collection of 10 static images of alphabet sign (A to J). Images took by us using smart mobile camera. Trained images consists only hand part which is explicitly cropped by us. Our tested database is a collection of 100 images (10 for each alphabet) taken by smart camera from 3 different persons. Limitation of our data is that the images are taken with only dark clothes.

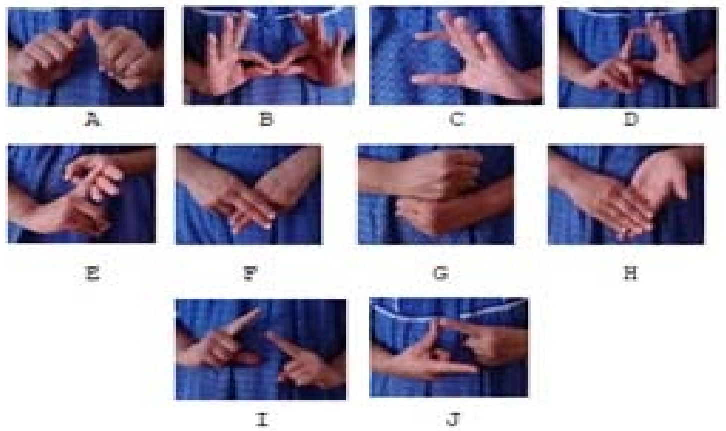

Fig 2. Training Dataset of Alphabets Sign A to J

3.3. Pre-Processing

Pre-processing is the first phase of proposed architecture. During this phase, testing database input the image for recognition. These images are converted into RGB images. These images further converted to gray scale image and then binary images. Next we use the skin color detection algorithm [21] for segmentation of hand part from the images. Further hand part is proceeding to next phase that is feature extraction.

3.4. Feature Extraction

Correlation-coefficient method is used to extract the similarity index value as a feature. Tested object is compares with the training objects and derive similar object from training dataset by identifies the nearest coordinates of these two objects. Next these closet coordinates are uses to normalize the correlation-coefficient. Then count the optimal pick value by calculating the correlation-coefficient values. This pick value is called as similarity index value. Correlation-coefficient algorithm is work based on following equation [22].

$$y(u,v) = \frac{\sum_{x,y}\left[f(x,y) - \bar{f}_{u,v}\right]\left[t(x-y, y-v) - \bar{t}\right]}{\left\{\sum_{x,y}\left[f(x,y) - \bar{f}_{u,v}\right]^2 \sum_{x,y}\left[t(x-y,y-v) - \bar{t}\right]^2\right\}^{0.5}}$$

Where,

- f is the image.
- $\bar{t}$ is the mean of the object
- $\bar{f}_{u,v}$ is the mean of f(x,y) in the region under the object
- x & y are the dataset (Here, x& y dataset are the training and testing data respectively)

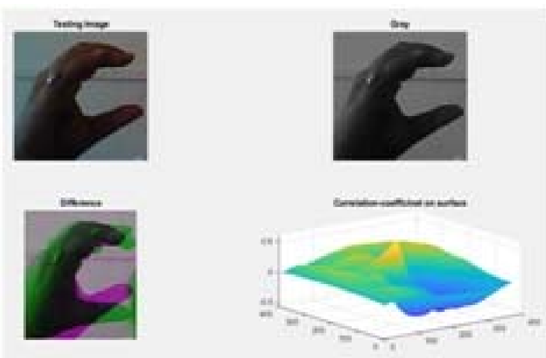

Fig 3. Compare testing and training image of alphabet C sign to find correlation-coefficient

Derived similarity index value and actual similarity index value of trained image were proceeding to next phase that is classification & recognition of sign.

3.5. SVM/ANN/HMM/NF Based Classification Algorithm

3.5.1 SVM
Support Vector Machine (SVM) is a machine learning algorithm on the base of supervised learning method [23]. Normally it is use for classification. In the SVM each feature values are put on the plot in n-direction. This plot is divided into two parts: 1. hyper plane and 2. Maximum Margin hyper plane. Soft margin and kernel functions are used for optimal utilization of SVM.

3.5.2 CNN
Convolutional neural network (CNN) algorithm is based on biological nerves system. In CNN [24], extracted features are introduced as neurons working in a unit to recognize word. Feed Forward neural Network is used to recognize sign and convert into word. Feed forward neural network consist of three layers, 1. Input 2. Hidden Layer 3. Output. Neurons are inputted and calculated with weights and further transfer to hidden layer. Next then aggregate function is applied at hidden layer and transfer the optimal solution to output layer.

3.5.3 HMM
Hidden Markov Model (HMM) [25] is utilized here for data extraction purpose for sign recognition. HMM is a collection of finite states of extracted features. Each state divides into two set of probabilities, a transition probability and a discrete or continuous output probability which further use as condition to find the final output. Training data is used by matching extracted features of testing data.

3.5.4 NF
Neuro-Fuzzy algorithm is the hybridization of fuzzy logic and inference rules where rules are compared the values of extracted features to classify the particular word [26].N-F work as following:

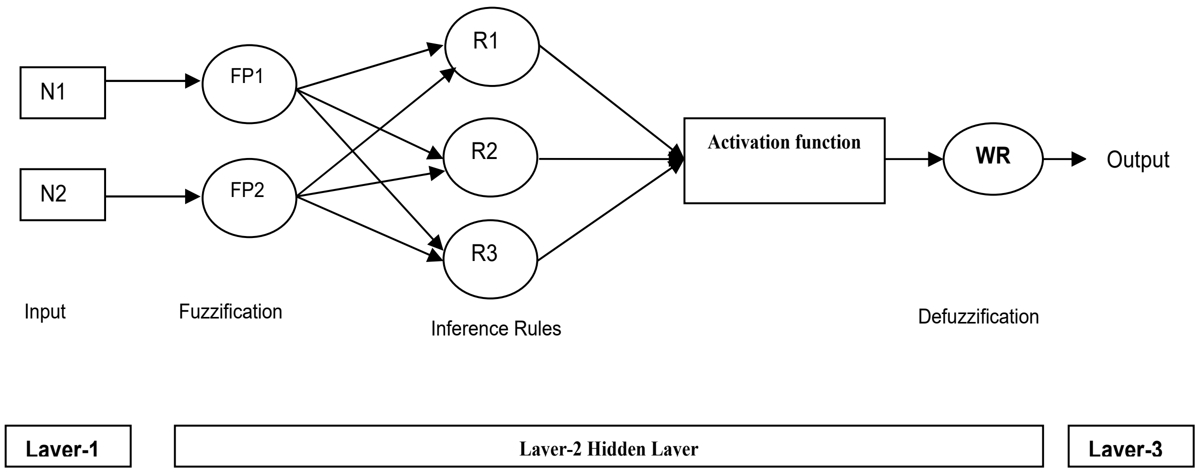

Fig 4. Graphical representation of Neuro-Fuzzy Algorithm

Above figure divide of three layers where first layer consists of a feature as neurons which have been continue to hidden layer (Fuzzification layer) and convert into crisp value. These crisp values further used inference rules and return the derived value in third layer.

4. Testing and Result Discussion

In our research we have trained 10 images (one for each alphabet A to J). We have tested 100 images by comparing features and other parameters with the trained images. Each trained and tested image is explicitly taken by mobile camera and used in .jpeg format. Proposed Image Processing techniques and Recognition algorithms are executed and took a stab at MATLAB. We have achieved 94% accuracy with the use of SVM, 70% accuracy with the use of HMM, 95% accuracy with the use of CNN and 97% accuracy with the use of NF algorithm. Following table shows average outcome of attempted pictures.

Hand Gestures for alphabets	Total Testing Images of True Positive classes	Evaluation Rate of SVM in %	Evaluation Rate of HMM in %	Evaluation Rate of CNN in %	Evaluation Rate of NF in %
A	10	100%	90%	100%	100%
B	10	100%	80%	100%	100%
C	10	90%	80%	90%	100%
D	10	100%	100%	100%	100%
E	10	100%	70%	100%	100%
F	10	90%	90%	80%	100%
G	10	90%	50%	90%	100%
H	10	90%	50%	100%	100%
I	10	80%	40%	90%	90%
J	10	100%	50%	100%	100%
Total	**100**	**94%**	**70%**	**95%**	**97%**

Table 2. Evaluation Results of SVM, HMM, CNN & NF algorithm

Following is the graphical representation of evaluation result.

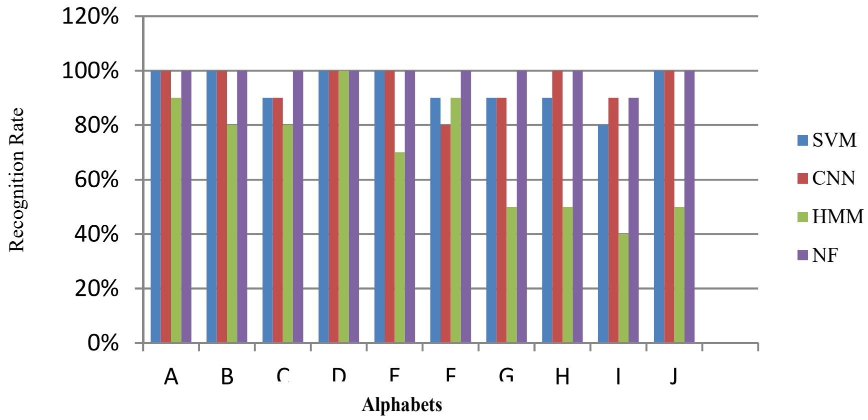

Fig 5. Recognition rates obtained by testing each method on sequences of gestures images alphabets A to J

5. Comparison of SVM, CNN, HMM & NF

5.1. Comparison of SVM, CNN, HMM & NF Based On Different Performance Evaluation Parameters

Performance Evaluation Parameters	SVM	CNN	HMM	NF
Time Complexity	It take both long training time & testing time	It take long training time and less testing time	It take both long training time & testing time	It take both less training time & testing time
Space Complexity	It consumes less memory	It consumes less memory	It consumes more memory then NF	It consumes less memory then other algorithms
Sample Complexity	Don't need large amount of training data	Needs large amount of training data	Needs large amount of training data	Needs large amount of training data
Parametercity	Require more parameters to classify images	Require less parameters to classify images	Require more parameters to classify images	Require less parameters to classify images then other algorithms
Response time	It take long time to give first response	It take less time to give first response	It take long time to give first response	It take less time to give first response

Table 3. Comparison of existing algorithms and Proposed Algorithm

Performance Evaluation Parameters	**SVM**	**CNN**	**HMM**	**NF**
Training Time Complexity	2.5 Seconds	1.5 Seconds	2.0 Seconds	0.9 Seconds
Testing Time Complexity	1.39 ms	1.08 ms	1.20 ms	0.89 ms
Space Complexity	500 bytes	480 bytes	490 bytes	445 bytes
Response time	1.05 ms	0.90 ms	1.15 ms	0.78 ms
Accuracy based on confusion matrix	93%	70%	92%	96%

Table 4. Comparison of existing algorithms and Proposed Algorithm based on evaluated values of different parameter

Confusion matrix

A confusion matrix is use to find the performance of classification model where matrix is evaluate by true positive and false positive class. To get more accuracy model has been evaluate by counting Precision, Recall & F-score. For the testing purpose we have taken 10 positive classes and 10 negative classes of each sign. Each class consists of 10 images of each sign respectively.

5.2 Advantages of Neuro-Fuzzy Approach Compare To SVM, CNN, HMM

5.2.1 Disadvantages of SVM, CNN, HMM

i. SVM doesn't perform well when we use large amount of data volume.
ii. CNN needs more feature symbols to evaluate optimal solution.
iii. CNN needs large amount of training data to find solution.
iv. HMM also requires large amount of training data.
v. HMM states require independent or annotated data volume of training set.

5.2.2 Advantages of Neuro-Fuzzy Approach

i. N-F consists merits of both individual algorithm neural network and fuzzy logic.
ii. N-F algorithm is described both self-learning and decision power.
iii. N-F faced fewer difficulties even though we use large amount of data volume.
iv. N-F needs less computational cost compare to other mention algorithms.

6. Conclusion

This paper conclude that Sign language recognition is the key feature since last few decades and very useful to the society as physically hearing-impaired or mute people can easily communicate in front of normal people. Recognition has been done using image processing and Classification. Classification has been performing on four algorithms: SVM, CNN, HMM and N-F which next describe the comparison among them in the form of evaluation results. On the base of evolution and comparison, conclusion is that, neuro-fuzzy network is the better classification algorithm for presented research-based system. N-F algorithms consist merits of both neural network and fuzzy logic makes it high computational algorithm compare to SVM, CNN, and HMM. We have achieved 94% accuracy with the use of SVM, 70% accuracy with the use of HMM, 95% accuracy with the use of CNN and 97% accuracy with the use of N-F algorithm. Paper describes a MATLAB editor which is used for implementation. Confusion matrix describe clear picture of accuracy of N-F approach compare to other algorithms.

7. Future Work

Though we have researched many things about sign language recognition, there are many techniques still apart from presented techniques for feature extraction as well as classification. One can also work on dynamic sign language recognition.

References

[1] Ashok Kumar Sahoo and Kiran Kumar, (2014) . MATLAB, Indian Sign Language Recognition using Skin Color Detection, International Journal of Applied Engineering Research, ISSN 0973-4562, Volume 9, Number 20 ,pp. 7347-7360.

[2] Madhuri Sharma, Ranjna Pal and Ashok Kumar Sahoo (2014). INDIAN SIGN LANGUAGE RECOGNITION USING NEURAL NETWORKS AND KNN CLASSIFIERS, ARPN Journal of Engineering and Applied Sciences, VOL. 9, NO. 8, AUGUST 2014, ISSN 1819-6608.
[3] Aliaa A. A. Youssif, Amal ElsayedAboutabl, Heba Hamdy Ali. (2011). Arabic Sign Language (ArSL) Recognition System Using HMM, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 11.
[4] L. Pigou et al. (2014). Sign Language Recognition Using Convolutional Neural Networks. European Conference on Computer Vision 6-12 September 2014.
[5] R. Sharma et al. (2013). Recognition of Single Handed Sign Language Gestures using Contour Tracing descriptor. Proceedings of the World Congress on Engineering 2013 Vol. II, WCE 2013, July 3 - 5, London, U.K.
[6] Y.F. Admasu, and K. Raimond. (2010). Ethiopian Sign Language Recognition Using Artificial Neural Network. 10th International Conference on Intelligent Systems Design and Applications, PP. 995-1000.
[7] P. V. V. Kishore and P. Rajesh Kumar (2012). A Video Based Indian Sign Language Recognition System (INSLR) Using Wavelet Transform and Fuzzy Logic, IACSIT International Journal of Engineering and Technology, Vol. 4, No. 5, October 2012.
[8] Gaolin Fang, Wen Gao, and Debin Zhao (2004). Large Vocabulary Sign Language Recognition Based on Fuzzy Decision Trees, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, VOL. 34, NO. 3, MAY 2004.
[9] K. Grobel and M. Assan (1997). Isolated sign language recognition using hidden Markov models, in Proc. Int. Conf. Systems, Man Cybernetics, pp. 162–167.
[10] B. Bauer and H. Hienz (2000). Relevant features for video-based continuous sign language recognition in Proc. 4th Int. Conf. Automatic Face Gesture Recognition, pp. 440–445.
[11] R. H. Liang and M. Ouhyoung (1998). A real-time continuous gesture recognition systemforsignlanguage in Proc.3rd Int.Conf. Automatic Face Gesture Recognition, pp. 558–565.
[12] T. Starner, J. Weaver, and A. Pentland (1998). Real-time American Sign Language recognition using desk and wearable computer based video, IEEE Trans. Pattern Anal. Machine Intell., vol. 20, pp. 1371–1375, Dec. 1998.
[13] SabahetaĐogić, Gunay Karl (2014). Sign Language Recognition using Neural Networks, TEM Journal – Volume 3 / Number 4 / 2014.
[14] Sajanraj T D and Beena M V. (2018). Indian Sign Language Numeral Recognition Using Region of Interest Convolutional Neural Network, 2nd International conference on Inventive Communication and computational technologies (ICICCT 2018).
[15] Sarfaraz Masood, Adhyan Srivastava, Harish Chandra Thuwal and MusheerAhmad (2018). Real-Time Sign Language Gesture (Word) Recognition from Video Sequences Using CNN and RNN, Springer Nature Singapore Pte Ltd. V. Bhateja et al. (eds.), Intelligent Engineering Informatics, Advances in Intelligent Systems and Computing 695.
[16] HoussemLahiani, Mohamed Elleuch, MonjiKherallah (2015). Real Time Hand Gesture Recognition System for Android Devices, 978-1-4673-8709-5/20 15 IEEE.
[17] Andres Jess´e Porfirio, Kelly La´ısWiggers, Luiz E. S. Oliveira, Daniel Weingaertner (2013). LIBRAS Sign Language Hand Configuration Recognition Based on 3D Meshes, 978-1-4799-0652-9/2013 IEEE.
[18] Hanji Wang, XiujuanChai, Xlinchen (2015). Sparse Observation (SO) Alignment for Sign Language Recognition, 0925-2312/2015 Elseiver.
[19] M.V. Davydov, I.V. Nikolski, V.V. Pasichnyk (2010). Realtime Ukrainian sign language recognition system, 978-1-4244-6585-9/2010 IEEE.
[20] WatcharinTangsuksant, Suchin Adhan, ChuchartPintavirooj (2015). American Sign Language Recognition by Using 3D Geometric Invariant Feature and ANN Classification, The 2014 Biomedical Engineering International Conference (BMEiCON-2014), 978-1-4799-6801-5/2014 IEEE.
[21] Hemina Bhavsar, Dr. Jeegar Trivedi (2019). Hand gesture recognition for Indian Sign Language using skin colour detection and Correlation-Coefficient Algorithm with Neuro-Fuzzy Approach, at IEEE International Conference on Advances in Computing, Communication and Control (ICAC3 19),December 20 - 21, 2019.
[22] Haralick, M., and Shapiro, L. (1992). Computer and Robot Vision, Volume II, Addison- Wesley, pp. 316-317.
[23] D. G. Lowe (2014). Distinctive image features from scale invariant key points, Int. J. Comput. Vis vol. 60, no. 2, pp. 91–110, Nov.
[24] G. R. S. Murthy, R. S. Jadon (2010). Hand Gesture Recognition using Neural Networks, IEEE Trans., Vol. 6.
[25] T. Starner& A. Pentland (1995). Real-time American sign language recognition from video using hidden markov models, Technical Report, M.I.T Media Laboratory Perceptual Computing Section, Technical Report No.375.
[26] Binh, N., Ejima, T. (2005). "Hand Gesture Recognition Using Fuzzy Neural Network", In: ICGST Conf. Graphics, Vision and Image Process (Cairo, 2005), pp. 1–6.

Authors Profile

MS. Hemina Bhavsar was born in 8th July 1986 in Valsad, Guajrat, India. She has completed Graduation in B.Sc. Computer Science and Post graduation in Master of Computer Science from the Veer Narmad South Gujarat university, Gujarat, India. Since 2017 she has perusing P.hd in the research area of Image processing and machine learning.

From 2010 to 2013 she has worked as Assistant Professor in Sardar Patel College, BCA, Navsari, Gujarat, India. Since 2013 she is working as Assistant Professor in S.S.Agrawal Institute of Computer Science, Navsari, Gujarat, India.

Dr. Jeegar Trivedi is working at Post Graduate Department of Computer Science & Technology, Sardar Patel University. He specializes in Artificial Intelligence especially in Neuro-Fuzzy Systems and Knowledge Based Intelligent Systems. Besides this his other interest areas are in field of Computer Networks & Software Engineering. He has also worked for U.G.C funded major research project and has also developed projects for software industries also. He has several
Publications in National/International Journals, Book Chapters and in proceedings of international conferences. He has delivered guest lectures and invited talks at several academic institutes and worked as resource person for different FDPs in his subject.