Recent Developments in Sign Language Recognition : A Review

Paranjape Ketki Vijay, Naphade Nilakshi Suhas, Chafekar Suparna Chandrashekhar &
Deshpande Ketaki Dhananjay

Vishwakarma Institute of Information Technology, Pune, Sr. No. 2/3/4, Kondhwa(Bk), Pune-48, India
E-mail : ketkiparanjape123@yahoo.co.in, neelakshinaphade@gmail.com, suparnachafekar@gmail.com &
ketakiddeshpande@gmail.com

***Abstract* - In the world of sign language, and gestures, a lot of research work has been done over the past three decades. This has brought about a gradual transition from isolated to continuous, and static to dynamic gesture recognition for operations on a limited vocabulary. In present scenario, human machine interactive systems facilitate communication between the deaf, and hearing people in real world situations.**

In order to improve the accuracy of recognition, many researchers have deployed methods such as HMM, Artificial Neural Networks, and Kinect platform. Effective algorithms for segmentation, classification, pattern matching and recognition have evolved. The main purpose of this paper is to analyze these methods and to effectively compare them, which will enable the reader to reach an optimal solution. This creates both, challenges and opportunities for sign language recognition related research.

***Keywords*- Sign Language Recognition, Hidden Markov Model, Artificial Neural Network, Kinect Platform, Fuzzy Logic.**

I. INTRODUCTION

In the present environment of interactive and intelligent computing, an efficient human–computer interaction assumes an utmost importance.

Gesture recognition has wide-ranging applications [1] such as the following:

- developing aids for the hearing impaired;
- enabling very young children to interact with computers; -recognizing sign language;
- medically monitoring patients
- distance learning/tele-teaching assistance;
- navigating and/or manipulating in virtual environments;

Most common forms of gesture recognition involve Sign Language Interpretation.

Sign Language Recognition implies conveying meaningful information through the use of hand gestures.

II. RELATED WORK

In the last two decades, several different methods of potential applications in the advanced gesture interfaces for HCI have been suggested but these differ from one to another in their models. Some of them are Neural Network, HMMs [2] and Fuzzy Systems [3].

The past decades have witnessed two specific patterns for categories of sign language recognition.

The first category is dependent upon electromechanical devices for example: The glove based systems. They can be used for measuring different parameters of gestures such as hand and finger position, angles, and tip recognitions etc. However, their inconvenience was a major problem with the signer, and they were declared as cumbersome.

The way the interaction between the signer and the system took place, was complicated and less natural. The second category revolved around the use of machine vision and image processing techniques to formulate vision based hand gesture recognition systems.

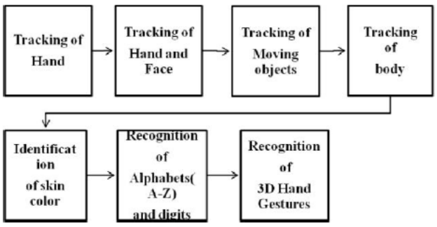

Tracking of Hand
Tracking of Hand and Face
Tracking of Moving objects
Tracking of body
Identification of skin color
Recognition of Alphabets(A-Z) and digits
Recognition of 3D Hand Gestures

Figure 1. Evolution of Sign Language Recognition

Ming, Huei[4] developed a system based on windowed template matching recognition strategy. It focuses on recognizing a continuous flow of signs in ASL, by enhancing the functionalities of the traditional data glove. A primitive signer dependent system was designed by Ouhyoung[5], which used Data glove, which solved the critical problem of end point detection. The system specifically used Taiwanese Lexical vocabulary. Starner and Pentland[6] presented a high accuracy computer visionbased method, which used Hidden Markov Models for American Sign Language recognition. They used the traditional dual camera method, one for tracking user's hand, and the other mounted on the desk. Caridakis and Diamenti[7] designed a robust system for consistent hand tracking and preservation of the resulting contours. Their scheme incorporated Markov chains, HMM, self-organizing maps for segmentation and feature extraction.

Lahamy and Litchi [9] developed a system for automatic gesture recognition using range cameras. Its ability tosimultaneously capture entire 3D point clouds was a preliminary step towards dynamic sign recognition. Chen and Georganas[10] developed a 3-D hand tracking and Analyzer system which used method based on Haar-like features and Adaboost algorithms based on stochastic parsing. This helped to achieve better accuracy against the complex, cluttered backgrounds. Nguyen and Shuichi[11] proposed a tracking, training and recognition ASL recognition system in an unconstrained environment. It is done using pseudo 2-D HMM (P 2-DHMM) and a Kalman filter and hand blob analysis. In the system developed by Zieren and Kraiss[12], Static areas are removed from the input video on pixel level and it gives moderate accuracy for person independent recognition.

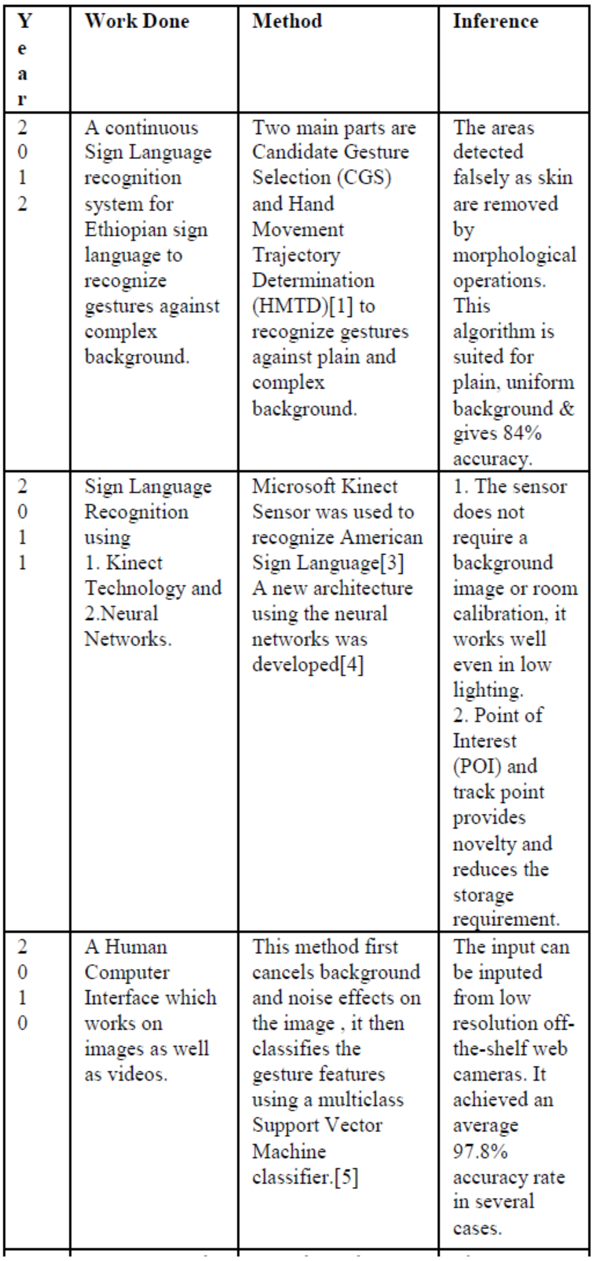
Peng and Wattanachote[13] addressed more practical and real-time problems through their approach that provided information retrieved from internet, allowing users to operate the system with his hand movements. It enabled the users to select functions through gestures, the action information for which, is reported by the system. Shreenivasan and Geethapriya[14] formulated a real-time system, which has applications in video games. Extraction and clustering of key points is done using rotation and scale invariant matching (RASIM) and k-harmonic means techniques. Ghosh and Ari [15] developed a system for human alternative and augmentative communication (HAAC).

A k-mean based radial basis function neural network (RBFNN) for classification was also proposed. Smith and Lobo[16] have presented a method to segment the hand over complex backgrounds such as the face, using Mixture of Gaussians (MOG) to detect the occluded region in the image. MacLean James[17] proposed the use of a back propagation neural network for recognition of gestures from a set of segmented hand images. This system showed promise in the field of language invariant teleconferencing.

Loeding and Sarkar[18] developed a system which made use of Signemes i.e. parts of signs that are present in most occurrences, taken from videos. Extaction of Signemes is done using Iterated Conditional Modes (ICM).

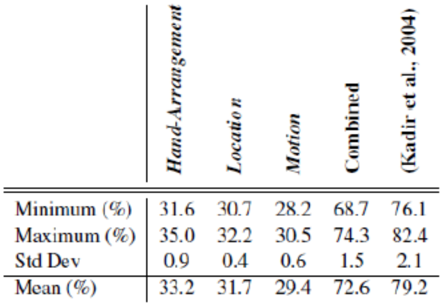
III. Recent trends

TABLE I. RECENT TRENDS IN SIGN LANGUAGE RECOGNITION



Year	Work Done	Method	Inference
2012	A continuous Sign Language recognition system for Ethiopian sign language to recognize gestures against complex background.	Two main parts are Candidate Gesture Selection (CGS) and Hand Movement Trajectory Determination (HMTD)[1] to recognize gestures against plain and complex background.	The areas detected falsely as skin are removed by morphological operations. This algorithm is suited for plain, uniform background & gives 84% accuracy.
2011	Sign Language Recognition using 1. Kinect Technology and 2.Neural Networks.	Microsoft Kinect Sensor was used to recognize American Sign Language[3] A new architecture using the neural networks was developed[4]	1. The sensor does not require a background image or room calibration, it works well even in low lighting. 2. Point of Interest (POI) and track point provides novelty and reduces the storage requirement.
2010	A Human Computer Interface which works on images as well as videos.	This method first cancels background and noise effects on the image , it then classifies the gesture features using a multiclass Support Vector Machine classifier.[5]	The input can be inputed from low resolution off-the-shelf web cameras. It achieved an average 97.8% accuracy rate in several cases.

2 0 0 9	A system used Hidden Markov Models, Boundary Tracing and Finger Tip Detection.	Boundary edge detection algorithm was used for Edge Detection, Clipping and Boundary Tracing.[4] HMMs were used to recognize gestures in real time.[5]	The accuracy obtained in this work had spelling errors resulting from the 5% error in the gesture recognition algorithm.



	Hand-Arrangement	Location	Motion	Combined	(Kadir et al., 2004)
Minimum (%)	31.6	30.7	28.2	68.7	76.1
Maximum (%)	35.0	32.2	30.5	74.3	82.4
Std Dev	0.9	0.4	0.6	1.5	2.1
Mean (%)	33.2	31.7	29.4	72.6	79.2

Fig: This fig shows an example of the extent to which the choice of a classifier affects the performance of an SLR syatem. Kladir model[30].

IV. EXISTING METHODOLOGIES

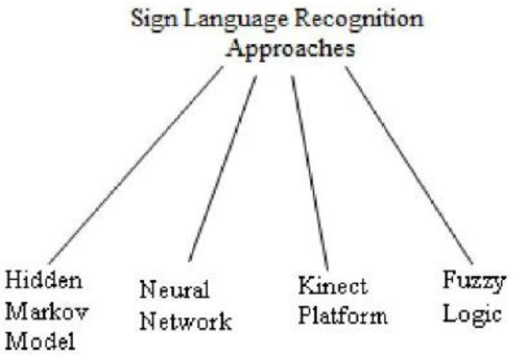


Figure 2. Recent Trends in Sign Language Recognition

A. Artificial Neural Networks

In recent times, Artificial Neural Networks are being perceived as the instrumental technology by choice for a lot many applications, for e.g. pattern recognition, speech recognition [19], prediction systems [20], identification and control of systems.

The Neural Network is an adaptive system that teaches itself to perform a function from the data sets. By adaptive, we mean, that the system parameters are subjective and may change during operations period, normally called the training phase. At the end of the training phase, the Neural Network parameters are defined and the system is deployed to resolve the problem, which constitutes our testing phase. It is basically a system based on the operation of biological neural networks, in other words, an emulation of biological neural system. In this, there are a set of input-output pairs, often provided by an external supervisors. On the basis of the difference between the desired response and the system output, error value is computed. This error information is fed back to the system and adjusts the system parameters in an organized fashion, which is termed as the 'System Learning Period'. The project implemented by Corneliu Lungociu [21] makes use of a web camera, which takes images of the signs made by the signer. These images are processed and the necessary characteristics for recognition are extracted and then used as inputs for an artificial neural network, which will recognize the sign.

The neural network used by them was a feed-forward network [22] with layered architecture. It consists of one input layer, one output layer and one hidden layer, making a total of three, with each layer fully connected with the following layer.

The number of neurons from the hidden layer is the mean between the number of neurons from the input layer and the number of neurons from the output layer. The aggregation function for each neuron is the weighted sum of its inputs and the transfer function is the sigmoid function [23], which is the basis for neuralic recognition. The network was trained in a supervised learning environment, with the most commonly used learning algorithm for neural networks with layered architectures, which is the back propagation algorithm [23], stochastic backpropagation, to be precise.

They successfully carried out the implementation to achieve a whopping 99% accuracy on a 36 imagery data set, with a system independent of the signer involved.

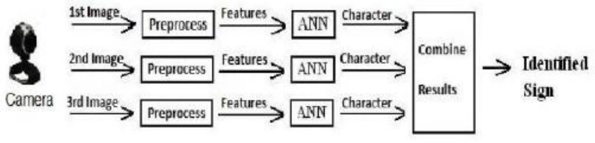


Figure 3. Recognition in Artificial Neural Networks

B. Hidden Markov Model: (HMM)

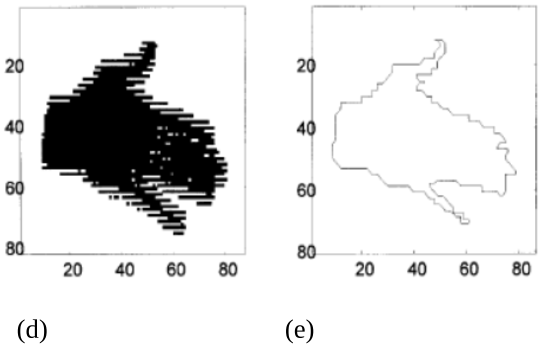
In case of dynamic process modeling, the approach to be chosen is stochastic. In nature, e.g. Hidden Markov Models (HMMs) [24] or Dynamic Bayesian Networks [25]. A timedomain process illustrates Markov property if the conditional probability density of an event, given all present and past events, depends only on the jth most recent event. If the current event depends solely on the most recent past event, then the process is termed a first order Markov process. It is considered as a useful assumption, while considering the orientations of hands of a signaler for time axis.

HMM is said to be a dual stochastic process that is looked upon by an underlying Markov chain with a definite number of states and a set of random functions, each associated with a single state. In consideration to discrete time, the process belongs to a state and gives an observatory sign according to the random function in accordance with current state. The transitions among the states have heuristically defined probabilities. For example, according to Mitra and Acharya:

1) Transition probability, which provides the probability for undergoing the transition;

2) Output probability, which defines the conditional probability of emitting an output symbol from a finite alphabet when given a state.

The HMM, known for its rich mathematical structure, is considered to be a widely used tool for efficiently modeling spatio–temporal information in the most natural way possible. The algorithms that can be employed in the pass include the Baum–Welch and Viterbi [17], for evaluation, learning, and decoding before the interpretation can actually be started upon The Hmm generalized topology is known as the 'Ergodic model', as per Acharya's term conditions, wherein any state can be reached from any other state.



(d) (e)

Fig. 4 (a-e) : an example of border extraction algorithm on a digital figure. Two hand shapes image and their contours obtained from the border extraction algorithm.

Generally, a 2-D projection is extracted from a 3-D skeletal model of a hand and experimentally, a set of input is taken out. The spatial component of the dynamic gesture is typically neglected, while the temporal component (having a start state, end state, and a set of observation sequences) is mapped through an HMM classifier with appropriate boundary conditions. A set of data are employed to train the classifier, and the test data are used for prediction verification.

The key issues as per the Lungociu experiment involve:

1) Evaluation: determining the of the sequence being generated by the model using Forward–Backward algorithm

2) Training or estimation: maximizing the probabilities using Baum–Welch algorithm

3) Decoding: recovering the state sequence by Viterbi algorithm

The state chart as given by Mitra:

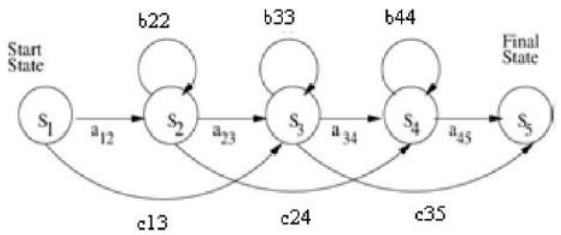


Figure 4. Five-state left-to-right HMM for gesture recognition

In the second approach to be discussed, which is the Yamato approach, a discrete HMM and a sequence of quantized (VQ)-labels have been used to recognize six classes of table tennis game strokes. The image, before actually allowing through the Hmm needs to be processed through several preprocessing steps such as low-pass morphological filtering to reduce the noise, background subtraction for precise extraction of the moving objects, and converting it to its binary version etc in order to generate blobs. The blobs roughly represent the actual gestures. These features are vector quantized, such that the image sequence becomes a sequence of VQ-labels, which are then processed by a discrete HMM.

HMM in sign language can broadly serve so many purposes such as word-level sign vocabulary, finger-spelling and non mechanical features [26]. However, in such a case the model fails to differentiate amongst undefined hand motions and the signer has to wear distinctly colored gloves on both hands, and maintain a physically immobile position (e.g. sitting in a chair) in front of the camera to aid hand tracking.

The HMM-based approach, however, requires extensive training sets for modeling. As mentioned above, the section discussed is all a part of the "the First subjective classification".

In the second stage of classification, each sign is modeled as an order 1 Markov chain, with each state in the chain representing a set of feature vectors generated from the classification in stage one. The Markov chain encodes temporal transitions of the hands. For dynamic gestures, each gesture is considered to be a sequence of epochs, where each epoch is characterized by a motion of distinct hand shapes and then Kalman filtering is used for hand tracking, and obtaining motion descriptors.

C. Kinect Technology

Kinect, code-named Project Natal, was developed by Microsoft and PrimeSense and released in November 2010. It is one of the latest technologies to join the list of globe shattering discoveries, in menial technological synagogue. Mainly targeted at consumers owning a Microsoft Xbox 360 console, it is the natural choice for building a framework for sign language recognition. The device itself features an RGB camera, a depth sensor and a multi array microphone, and is capable of tracking users' bodily movements.

After Kinect was released, Adafruit Industries which is known for open source hardware electronics released the completely documented drivers for Kinect. These are under an open source license. It includes a feature-rich open source framework that can be combined with closed source middleware called NITE for skeletal tracking and recognition of hand gestures. The Kinect demo of CopyCat also shows high accuracy using skeletal tracking even without regarding different handshapes. The goal of this work is to implement a framework for recognition of isolated signs.

The framework[27] created as part of this work by Simon Lang was called Dragonfly- Draw gestures on the fly, and mainly consists of two classes that users could use in assimilation with their software. The Depth camera serves as an interface to OpenNI, i. e. it updates the camera image and reports the data of the skeleton joints of the body parts, and the latter processes the data set for recognition.

The framework [27] uses Positive negative worst rate ratio algorithm for recognition, along with k clustering and classification. A Depth Camera supplied the framework with data of the users' joint positions in the form of threedimensional vectors which contain the coordinates of different body parts, e. g. right elbow, left hand or head. They can be processed and then serve as observations, thus offering 3D data of the environment without a complicated camera setup and dimly lit surroundings.

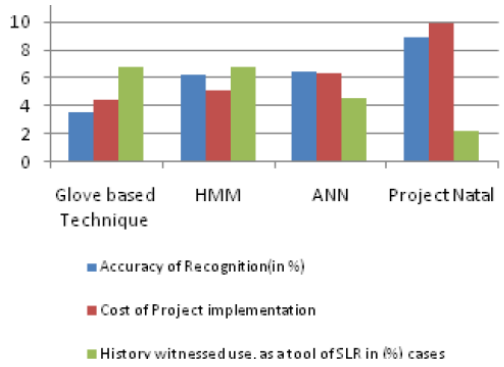


Fig: This is a representation of our survey, combining the overall project estimate, accuracy and trend belief over the years for an overview.

V. CONCLUSION

After the survey on the approaches used in various vocabulary-based sign language recognition systems, we can give an opinion about the methodologies and algorithms involved. Most of the times, a combination of different methods and algorithms has to be used to achieve a moderate to acceptable rate of recognition. For example some methods are suitable only against dark backgrounds. A system which gives maximum efficiency, has low cost, and is an optimal mixture of methods, giving results against complex backgrounds as well, should be preferred. From a technical point of view, there is a vast scope in future for research and implementation in this very field. The upcoming years could witness a combinatorial explosion of different methodologies, such as using several HMMs in parallel, independent or coupled usage of ANN and HMMs etc. The ultimate gain of the proposed study is enormous.

REFERENCES

[1] Acharya Tinku, Mitra Sushmita, “Gesture Recognition: A Survey” in *Ieee Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews, Vol. 37, No. 3, May 2007.*

[2] Czapnik Karol, Kasprazak WAlodzimierz, Wilkowski Arthur, “Hand Gesture Recognition in Image equences Using Active Contours and HMMs.”

[3] D. M. Gavrila, “The visual analysis of human movement: A survey.”

[4] Rung-Huei Liang Ming Ouhyoung, "A Real-time Continuous Alphabetic Sign Language to Speech Conversion VR System" in *Communications & Multimedia Lab., Computer Science and Information Engineering Dept, National Taiwan University, Taipei, Taiwan.*

[5] Rung-Huei Liang Ming Ouhyoung, "A Real-time Continuous Gesture Recognition System for Sign Language" in *National Taiwan University, Taipei, Taiwan.*

[6] Pentland Alex, Starner Thad, Weaver Joshua, "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video" in *Ieee Transactions On Pattern Analysis And Machine Intelligence, Vol. 20, No. 12, December 1998.*

[7] Caridakis George, Diamanti Olga et. al., "Automatic Sign Language Recognition: vision based feature extraction and probabilistic recognition scheme from multiple cues" in *Speech Communication and Signal Processing Group National Technical Conference, Athens.*

[8] Joshua R., "A Method for Hand Gesture Recognition." in *ACM Mid-Southeast Chapter Fall Conference, 2002.*

[9] Lahamy Herve, Litchi Derek, "Real-Time Hand Gesture Recognition Using Range Cameras."

[10] Chen Qing, Georganas Nicholas D, Petriu "Real-time Vision-based Hand Gesture Recognition Using Haar-like Features" in *Instrumentation and Measurement Technology Conference – IMTC 2007 Warsaw, Poland, May 1-3, 2007.*

[11] Nguyen Dang Binh, Enokida Shuichi, Toshiaki Ejima, "Real-Time Hand Tracking and Gesture Recognition System" *in GVIP 05 Conference,19-21 December 2005, CICC, Cairo, Egypt.*

[12] Kraiss, "Recent developments in visual sign language recognition." [13] Peng Sheng-Yu, Wattanachote Kanoksak, "A Real-Time Hand Gesture Recognition System for Daily Information Retrieval from Internet" in *Fourth International Conference on Ubi-Media Computing*.

[14] Geethapriya J and Srinivasan A, "A New Framework for Real-time Hand *Gesture Detection and Recognition" in International Conference on Information and Network Technology (ICINT 2012), IPCSIT vol. 37 (2012) © (2012) IACSIT Press, Singapore 2012.*

[15] Ghosh, D.K. , Ari, S. *"A Static Hand Gesture Recognition Algorithm Using K-Mean Based Radial Basis Function Neural Network" in Information, Communications and Signal Processing (ICICS) 2011 8th International Conference on* 13-16 Dec. 2011.

[16]. Smith Paul, Lobo Niels-da-Vitori, "Resolving Hand over Face occlusion."

[17] James MacLean, "Fast Hand Gesture Recognition for Real-Time Teleconferencing Applications".

[18]. Loeding Barbara, Sarkar Sudeep, "Automated Extraction Of Signs From Continuous Sign Language Sentences Using Iterated Condition Modes."

[19] K. R. Linstrom and A.J. Boye, "A neural network prediction model for a psychiatric application" in *International Conference on Computational Intelligence and Multimedia Applications, pp. 36-40, 2005.*

[20] M. D. Skowronski and J.G. Harris, "Automatic speech recognition using a predictive echo state network classfier" in *Neural Networks, Volume 20, Issue 3, pp:414-423, April 2007.*

[21] Lungociu Corneliu, "Real Time Sign Language Recognition Using Artificial Neural Networks." in *Studia Univ. Babes_Bolyai, Informatica, Volume Lvi, Number 4, 2011.*

[22] Peter Wray Vamplew , PhD Thesis: "Recognition of Sign Language Using Neural Networks" for Flinders University of South Australia, 1990.

[23] Tom M. Mitchell, "Machine Learning" in *McGraw-Hill, 1997*.

[24] K. R. Linstrom and A.J. Boye. "A neural network prediction model for a psychiatric application" in *International Conference on Computational Intelligence and Multimedia Applications, pp. 36-40, 2005.*

[25] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in timesequential images using hidden Markov model," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn., Champaign, IL, 1992, pp. 379–385.*

[26] A. J. Viterbi, "Error bounds for convolution codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory, vol. 13, no*. 2, pp. 260–269, Apr. 1967.

[27] R. Bowden, A. Zisserman, T. Kadir, and M. Brady, "Vision based interpretation of natural sign languages," in Proc. 3rd Int. Conf. Comput. Vis. Syst., Graz, Austria, Apr. 2003.

[28] Lang Simon, "Sign Language Recognition with Kinect".

[29] Mekala,P, "Real-time Sign Language Recognition based on Neural Network Architecture" in System Theory (SSST), 2011 IEEE 43rd Southeastern Symposium 14-16 March 2011.

[30] Helen Cooper, "Sign Language Recognition using Sub-Units" in *Journal of Machine Learning Research 13 (2012) 2205-2231.*

◈◈◈