Combine Approach for Speech and Gesture Recognition

Nutan D Sonwane
Department of Computer science & Engineering
G.H.Raisoni College of engineering, Nagpur
India

Prof.Sharda Chhabria
Department of Information Technology
G.H.Raisoni College of engineering, Nagpur
India

Dr.R.V.Dharaskar
Director of Matoshri Pratishthan's Group of Institution, MPGI Integrated campus, Nanded
India

ABSTRACT
Gesture and Speech based human Computer interaction is attractive attention across various areas such as pattern recognition, computer vision. Thus kind of research areas find many kind of application in Multimodal HCI, Robotics control, Sign language recognition. This paper presents head and hand Gesture as well as Speech recognition system for human computer interaction (HCI).This kind of vision based system can show the capability of computer. Which understand and responding to the hand and head gesture also for Speech in form of sentence. This recognition system consists of two main modules namely 1.Gesture recognition 2.Speech recognition, Gesture recognition consists of various phases. i. image capturing, ii. Feature extraction of gesture iii. Gesture modeling (Direction, Position, generalized), 2.Speech recognition consists of various phases i. taking voice signals ii. Spectral coding iii. Unit matching (BMU) iv. Lexical decoding v. syntactic, semantic analysis. Compared with many existing algorithms for gesture and speech recognition.

General Terms
Pattern recognition, Human computer interaction, Speech recognition, Gesture recognition.

Keywords
Self organizing map, Best matching unit, Hidden morkov model, Principal component analysis.

1. INTRODUCTION
The architecture for gesture recognition, fusing separate component model all of which are based on hand trajectory. The approach involves a combination of Self Organizing Maps and Markov Models for gesture trajectory classification, using the trajectory of the hand segment and direction of motion during a gesture. This classification scheme is based on the transformation of a gesture representation from series of coordinates and movements to a symbolic form and building probabilistic models based on these transformed representations. A self organizing map is proposed for recognition of objects based on tactile shape perception. The data in a speech recognition system, Training takes as input a large number of speech utterances along with their transcriptions into phonemes and outputs the speech models for the phonemes. The utterances to be recognized a spectral analysis stage, also called the feature extraction stage. Typical feature representations are smoothed. Spectra or linear Prediction. Automatic speech recognition is a process by which a machine identifies speech. The machine takes a human utterance as an input and returns a string of words phrases or continuous speech in the form of text as output. The conventional method of speech recognition insists in representing each Word by its feature vector and pattern matching with the statistically available vectors. Its most effective application is the development of strong and friendly interfaces for human-machine interaction, since gesture and speech is a natural and powerful way of communication.

2. LITERATURE REVIEW
A body of the literature survey suggests that people naturally tend to do activity with which they interact. It also observes how people use speech and gesture when interacting with system.

2.1 Principle Component Analysis: Solomon Raju Kota, J.L.Raheja, Ashutosh Gupta describe a method for gesture recognition It is a classical feature extraction technique widely used in the field of pattern recognition and computer vision [1]. The gesture recognition using PCA algorithm that involves two phases

• Training Phase • Recognition Phase: During the training phase, each gesture is represented as a column vector, with each entry corresponding to gesture pixel. These gesture vectors are then normalized with respect to average gesture.

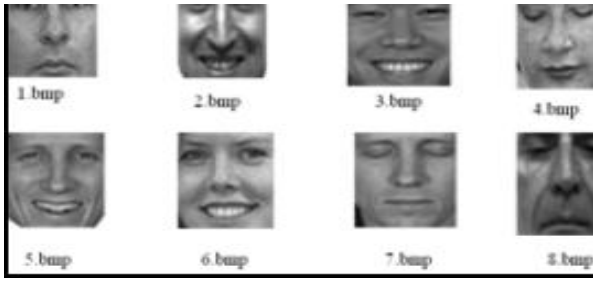


Fig.1: Database gestures

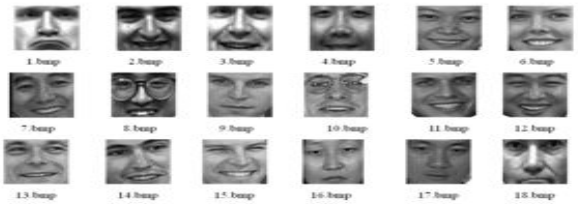


Fig.2: Test database gestures

2.2 Support Vector Machines: It is a classical statistical technique for analyzing the covariance structure of multivariate data.

2.3 Hidden Markov Model
Hidden Markov model is one of the major methods to recognize gestures in computer vision and pattern recognition. Hidden Markov models can easily handle simple gesture recognition, but not efficient enough for complicated gesture. Dynamic Bayesian networks with the topology of HMMs have richer information representation than HMM [4].

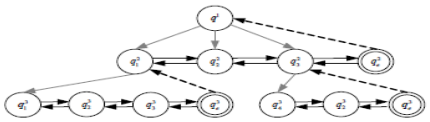


Fig..3: The representation of the HMM for recognizing hand gesture activities.

2.4 Self-Growing and Self-Organized Neural Gas (SGONG) network.

E. Stergiopoulou and N. Papamarkos [2] describe a method which is an unsupervised neural classifier. It achieves clustering of the input data, so as the distance of the data items within the same class (intra-cluster variance) is small and the distance of the data items stemming from different classes (inter-cluster variance) is large. Moreover, the final number of classes is determined by the SGONG during the learning process. It is an innovative neural network that combines the advantages both of the Kohonen Self-Organized Feature Map (SOFM) and the Growing Neural Gas (GNG) neural classifiers. The SGONG consists of two layers, i.e. the input and the output layer Segmented image

Fig.4: SGONG network (a) Original image (b)

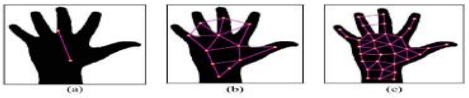


Fig..5: Growth of the SGONG network: (a) starting

2.5 Spatio-Temporal Feature-Extraction Techniques

Presents various spatio-temporal feature extraction techniques [5] with applications to online and offline recognitions of isolated Arabic Sign Language gestures. The temporal features of a video-based gesture are extracted through forward, backward, and bidirectional predictions. The prediction errors are threshold and accumulated into one image that represents the motion of the sequence. The motion representation is then followed by spatial-domain feature extractions. As such, the temporal dependencies are eliminated and the whole video sequence is represented by a few coefficients.

2.6 VLSI for 5000-Word Continuous Speech Recognition

Young-kyu Choi and Kisun You, Jungwook Choi, and Wonyong Sung describe a method a VLSI chip for 5,000 word speaker-independent continuous speech recognition. This chip employs a context-dependent HMM (hidden Markov model) based speech recognition algorithm, and contains emission probability and Viterbi beam search pipelined hardware units. The feature vector for speech recognition is computed using a host processor in software in order to adopt various enhancement algorithms.

3. SYSTEM ARCHITECTURE

Gesture recognition is the process by which gesture made by the user are known to the system. Gestures components are the Head and hand poses. Gestures are recognized using rule-based system according to predefined model with the combinations of the pose classification results of three segments at a particular image frame.

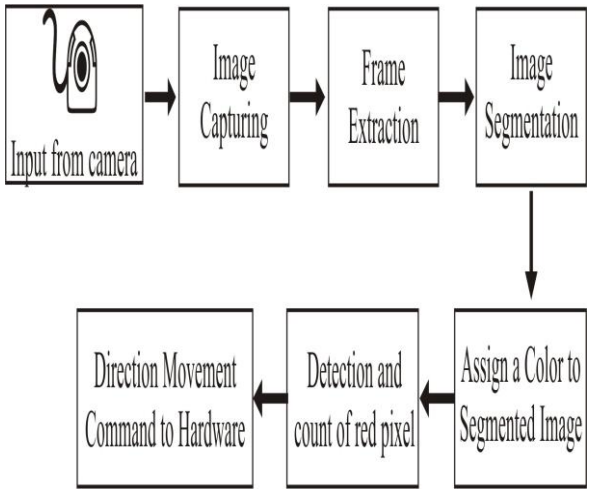
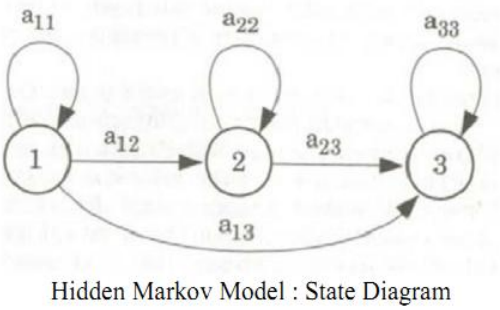


Fig: 6 Gesture recognition

3.1 SPEECH RECOGNITION

It is the ability of a computer to recognize general, naturally flowing utterances from a wide variety of users. Automatic speech recognition is a process by which a machine identifies speech. The machine takes a human utterance as an input and returns a string of words, phrases or continuous speech in the form of text as output. The conventional method of speech recognition insists in representing each word by its feature vector and pattern matching with the statistically available vectors.



Hidden Markov Model : State Diagram

Fig. 7 Speech recognition

A hidden Markov model can be used to model an unknown process that produces a sequence of observable outputs at discrete intervals where the outputs are members of some finite alphabet. These models are called "hidden" Markov models precisely because the state sequence that produced the observable output is not known-it's "hidden." HMMs have been found to be especially way for modeling speech processes. [3]

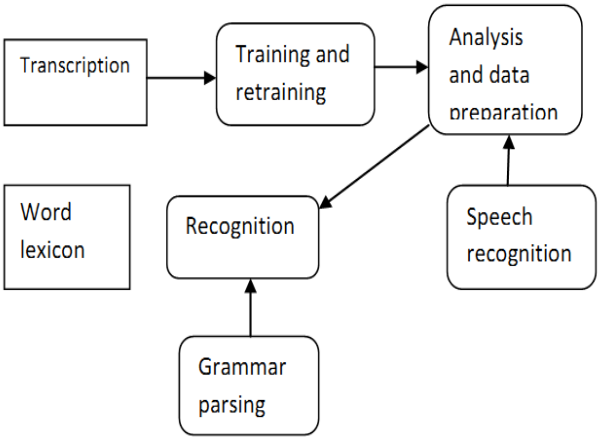


Fig.8 Speech recognition

3.2 Self organizing map

A self-organizing map consists of components called nodes or neurons. Associated with each node is a weight vector of the same dimension as the input data vectors and a position in the map space. The usual arrangement of nodes is a regular spacing in a hexagonal or rectangular grid. The self-organizing map describes a mapping from a higher dimensional input space to a lower dimensional map space.

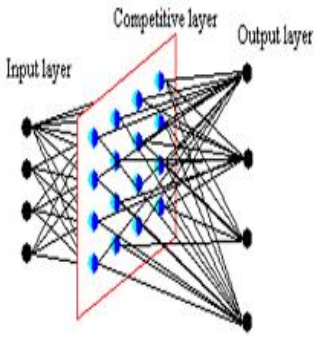


Fig 9.SOM network

3.2.1 Input Layer

Accepts multidimensional input pattern from the environment. An input pattern is represented by a vector. Each neurode in the input layer represents one dimension of the input pattern. An input neurode distributes its assigned element of the input vector to the competitive layer.

3.2.2 Competitive layer

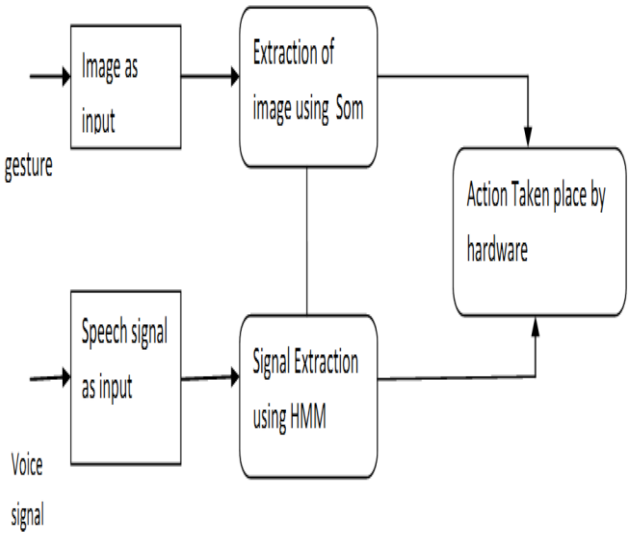
Each neurode in the competitive layer receives a sum of weighted inputs from the input layer. Every neurode in the competitive layer is associated with a collection of other neurodes which make up its 'neighborhood'. We can organize Competitive layer on any dimension. Upon receipt of a given input, some of the neurodes will be sufficiently excited to fire. This event can have either an inhibitory, or an excitatory effect on its neighborhood. The model has been copied from biological systems, and is known as 'on-center, off-surround' architecture, also known as lateral feedback / inhibition.

3.2.*3 Output layer*

Organization of the output layer is application-dependent. Strictly speaking, not necessary for proper functioning of a Kohonen network. The "output" of the network is the way we choose to view the interconnections between nodes in the competitive layer. If nodes are arranged along a single dimension, output can be seen as a continuum.

4. CONCLUSION

A gesture is a motion of the body that conveys information; in this paper, we focus on hand gestures and information conveyed from these gestures. And also on the speech and information conveyed from speech. Gesture taxonomy can be formalized in a scaling continuum: gesticulation, speech-linked, pantomime, emblems and sign languages.



5. REFERENCES

[1] Solomon Raju Kota,J.L Reheja,Ashutosh Gupta,Archna Rathi, Shashikant Sharma"Principal component analysis for Gesture recognition using systemC"2009 international Conferences in advance technology in communication and Computing 2009IEEE

[2] E. Stergiopoulou and N. Papamarkos "A New Technique for Hand Gesture Recognition" 1-4244-0481-9/06/ © 2006 IEEE

[3] Tan Wenjun, Wu Chengdong, Zhao Shuying Jiang Li "Dynamic Hand Gesture Recognition Using Motion Trajectories and Key Frames"2010

[4] G.R.S Murthy, R.S Jadon "Hand gesture recognition using neural network" in 2nd International Advance Computing Conference 2010

[5] Tamer Shanableh, Khaled Assaleh, Senior Member, IEEE, and M. Al-Rousan "Spatio-Temporal Feature-Extraction Techniques for Isolated Gesture Recognition in Arabic Sign Language". 1083-4419/ © 2007 IEEE

[6] Yean Choon Ham, Yu Shi "Developing a Smart Camera for Gesture Recognition in HCI Applications" The 13th IEEE International Symposium on Consumer Electronics (ISCE2009) 978-1-4244-2976-9/09/$25.00 2009

[7] Sivalogeswaran Ratnasingam, T. M. McGinnity "Object Recognition Based on Tactile Form Perception" in IEEE 2011

[8] Anjali Kalra, Sarbjeet Singh, Sukhvinder Singh "Speech Recognition" International Journal of Computer Science and Network Security, VOL.10, 2010

[9] George Caridakis, Kostas Karpouzis, Athanasios Drosopoulos, Stefanos Kollias" SOMM: Self organizing Markov map for gesture recognition" Pattern Recognition Letters 31, 2010

[10] Jagdish Lal Raheja, Radhey shyam "Real Time Robotic Hand Control Using Hand Gesture" 978-0-7695-3977- 5/10 © 2010 IEEE

[11] Mr. Chetan A. Burande, Prof. Raju M. Tugnayat, Prof.D. Nitin K. Choudhary "Advanced Recognition Techniques for Human Computer Interaction." 978- 1-4244-5586-7/10. 2010 IEEE