International Journal of Electronics Communication and Computer Engineering
Volume 5, Issue 2, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Improved ASL based Gesture Recognition using HMM for System Application

Shalini Anand
Research Scholar, RCEW, Jaipur

Manju Mathur
Assistant Professor, RCEW, Jaipur

Dinesh Goyal
Associate Professor, SGVU, Jaipur

***Abstract* – Gesture recognition is a growing field of research and among various human computer interactions; hand gesture recognition is very popular for interacting between human and machines. It is non verbal way of communication and this research area is full of innovative approaches. This project aims at recognizing 34 basic static hand gestures based on American Sign Language (ASL) including alphabets as well as numbers (0 to 9). In this project we have not considered two alphabets i.e J and Z as our project aims as recognizing static hand gesture but according to ASL they are considered as dynamic. The main features used are optimization of the database using neural network and Hidden Markov Model (HMM). That is the algorithm is based on shape based features by keeping in the mind that shape of human hand is same for all human beings except in some situations.**

***Keywords* – American Sign Language (ASL), Hidden Markov Model (HMM), Gesture Recognition.**

I. INTRODUCTION

In recent time with the advancement in the pattern recognition field and related computer vision, hand gesture recognition has turned vital form of non verbal communication. It has gained much importance also due to the fact that use of computer for any lay man has to be user friendly so that computer literacy does not hamper the usability of computer.

With the need of easy interaction between human and computer, a lot has been done to improvise/replace the age old input devices like keyboards and mouse.

Many researches are being done to make computers understand human language and develop human computer interfaces (HCI).

Gesture recognition is a field, in which we have large number of innovations. Gestures can be defined as any action taken physically to pass or symbolize information.

Gesture Analysis

It is not a good idea to carry with particular useful definition of gestures because of its wide variety of applications varying all over and a valid logic can only specify a particular domain of gestures. Many people have defined the definition of gestures in their own terms, so the definition is still arbitrary in wide array. In a lay man definition we can define gestures as a physical activity that conveys some message or information, be it can be facial expressions, hand movements, body language etc. Gesture can be defined as the motion of the body in order to communicate with others [2], and in order to complete a successful communication between sender and receiver, they must possess same set of gestures or should understand the meaning being conveyed. We can broadly classify gestures in two types: static and dynamic. Here static gestures mean static with respect to time. For example- the stop sign signal is static gesture while when we wave our hand when we move away means goodbye- is an example of dynamic gesture.

In gesture analysis, there are mainly three steps to follow

1. Hand localization
2. Hand feature extraction
3. Hand model parameter computation of features

There are two types of classification

(i) Rule Based Classification
(ii) Learning Based Classification

Rule-based approaches use a set of manually encoded rules between feature inputs. Given an input gesture a set of features are extracted and compared to the encoded rules, the rule that matches the input is outputted as the gesture. Human hand can be used as an input device for human-computer interaction and using hand as a mode of interaction is one of the promising methods in comparison to other input devices like mouse, keyboard. Human gesture recognition includes mainly three steps:

1. Image segmentation (pre-processing)
2. Feature extraction
3. Classification

The most important factor in human gesture recognition is the selection of good features. In this chapter we will be discussing the review of hand gesture recognition methods.

New researches in the field of hand gesture recognition [9-12] proved its importance in human computer interaction (HCI.

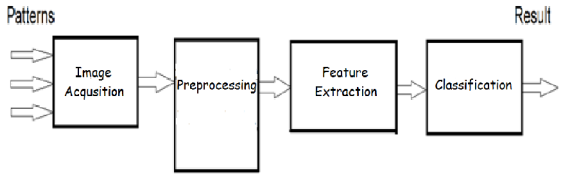
II. AMERICAN SIGN LANGUAGE

American Sign Language (ASL) is the predominant sign language of deaf communities in the United States and English-speaking parts of Canada. ASL originated in the early 19th century in the American School for the Deaf (ASD) in Hartford, Connecticut, from a situation of language contact. Since then, ASL use propagated widely via schools for the deaf and deaf community organizations. Despite its wide use, no accurate count of ASL users has been taken, though reliable estimates for American ASL users range from 250,000 to 500,000 persons, including a number of children of deaf adults.

Star-ner and Pentland developed a hand glove environment system which is capable of recognizing 40 characters form the American Sign Language (ASL) [5]. They have performed 2 experiements with the word accuracy of 99 and 92 percent and both the experiements were applied on 40 characters of American Sign Language. In first experiement they use hand wearing

coloured gloves with the background in contrast and in the second experiement only hand was traced without any gloves hence, the less accuracy in result. Their system uses one colour camera which will track hands in real-time scenario and will predict characters from the database of American Sign Language using Hidden Markov Models (HMM's) [15].

Etsuko Ueda and Yoshio Matsumoto [23] gave a technique for hand pose estimation that can be used for vision-based human interfaces. In this method we have to estimate all the joint angles to access objects in the virtual world. Multi-viewpoint camera system takes multiple images which are used to extract the hand regions. These multi-viewpoint images can be processed to reconstruct the hand pose as a "voxel model" [23]. Further all joint angles can be estimated using three dimensional model fitting between hand model and voxel model [23]. In 2011 Sanjay Meena had given a technique which is capable of recognizing characters of American Sign Language with the accuracy level of 94.6%, this was basically a training technique that works on linear classification and Support Vector Machine (SVM) [68].



After segmentation and feature extraction, recognition or classification makes up the last step of the process of hand gesture recognition. Hand gesture classification can be done in two ways as explained [23].

1. Rule based approach
2. Machine learning based approach

In rule based approaches, the rule need to be created manually on the basis of feature vectors, and the ones matching the rule can be recognised as the final result. But, this involves human handling capacity of creating so many rules and that is the disadvantage of this approach as it is very difficult to create a vast variety of rules manually which is required to get perfect recognition .

Some of the machine based learning classification methods are discussed below:

III. Neural Networks

This method is based on modeling of the neuron, human nervous system element and its interaction with other neurons for information transfer. Each node consists of an input function which computes the weighted sum and an activation function that generates the response based on the weighted sum. There are two types of neural networks-

1. Feed forward
2. Recurrent

IV. Hidden Markov Model

The Hidden Markov Model (HMM) classifiers hold a place in the class of trainable classifiers. HMM classifiers represents a statistical model, in which the most probable matching gesture class is determined for a given feature vector, based on the training data. It is one of the widely exploited methods for temporal gesture recognition.

An HMM consists of states and state transitions with observation probabilities. A separate HMM need to be trained to watch the gesture and the recognition of the gesture is based on the generation of maximum probability by a particular HMM. It does require considerable training effort and has complex working nature; the results are often unpredicted because of the hidden nature. Baum-Welch re-estimation algorithm, which adapts the internal states of the HMM according to some feedback concerning the accuracy was used to train the HMM.

$$= \frac{\sum_{d=1}^{D}\sum_{t=1}^{T} P(z_t^{(d)} = i|X^{(d)};\theta^s) I(x_t^{(d)} = j)}{\sum_{d=1}^{D}\sum_{t=1}^{T} P(z_t^{(d)} = i|X^{(d)};\theta^s)}$$

V. Applications of Gesture Recognition

In this section we will discuss the applications of hand gesture in various fields.Various applications have been used for hand postures and gestures as alternative level of interaction in different application domains, as mentioned in including virtual environments, smart surveillance, sign language translation, medical systems etc. This section gives a brief overview of some gesture recognition application areas.

5.1 Sign language recognition

Sign languages are the most raw and natural form of languages could be dated back to as early as the advent of the human civilization, when the first theories of sign languages appeared in history. It has started even before the emergence of spoken languages. Since then the sign language has evolved and been adopted as an integral part of our day to day communication process. [23]. It can be used for disabled peopled when communicating with the other people, and with the computer as well [23]. American Sign Language in [31, 03] is one example that has received utmost attention in the gesture literature.

5.2 Robotics, human manipulation

One of the effective applications that can utilize hand postures and gestures is robot tele-manipulation [29]. Tele-robotic applications are typically classified under space exploration and military research domain.

5.3 Gesture to speech

Gesture-to-speech application which converts hand gestures into speech, this system enables hearing-impaired people to communicate with their surrounding environments through computers and interacts easily with other people even without the knowing for the sign language [29].

5.4 Games

For computer games, [36] applied gesture recognition on virtual game applications.[37] Used hierarchical recognition of human gestures for sports video annimation..

International Journal of Electronics Communication and Computer Engineering
Volume 5, Issue 2, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

5.5 Television control

Last application for hand postures and gestures is controlling Television devices [29].Developed a system to control a television set by hand gestures. Using an open hand and the user can change the channel, turn the television on and off, increase and decrease the volume, and mute the sound.

5.6 Tele Presence

Tele presence is that area of technical intelligence which aims to provide physical operation support that maps the operator arm to the robotic arm to carry out the necessary task. The prospects of tele presence includes space, undersea mission, medicine manufacturing and in maintenance of nuclear power reactors.

VI. Optimizing the Image

In this step the image format is changed from jpg to pgm (portable graymap format) which is given as input by the user for pattern matching.PGM mage is of netpbm source. The phrase Netpbm format commonly refers to any or all of the members of a set of closely related graphics formats used and defined by the Netpbm project. The portable pixmap format (PPM), the portable graymap format (PGM) and the portable bitmap format (PBM) are image file formats designed to be easily exchanged between platforms.

In 1993, the Netpbm library was developed to replace the unmaintained Pbmplus. It was simply a repackaging of Pbmplus with additions and fixes submitted by people all over the world.

File format description:-

Each format differs in what colors it is designed to represent:

- PBM is for bitmaps (black and white, no grays)
- PGM is for grayscale
- PPM is for "pixmaps" which represent full RGB color.

PGM example

The PGM and PPM formats (both ASCII and binary versions) have an additional parameter for the maximum value (numbers of grey between black and white) after the X and Y dimensions and before the actual pixel data. Black is 0 and max value is white. There is a newline character at the end of each line.

Example (magnified)

VII. Proposed Work

The proposed algorithm has two main stages:-

1. Stage 1: Optimization of data
2. Stage 2: Recognition of gesture

Explanation of Algorithms:

STAGE I: It is further classified into three steps they are as mentioned

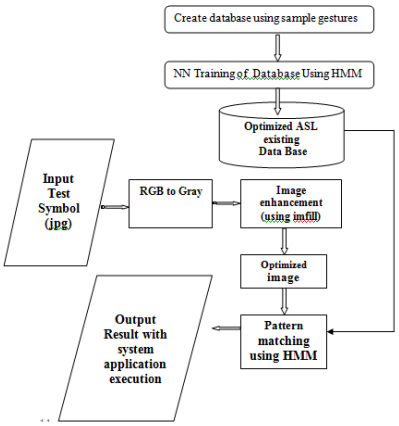


Fig.3.1 Flow Chart for Proposed Methodology

(i) Create database using sample gestures: As per the American Sign Language (ASL) database has been created and collected at one place. There are three types of gesture recognition systems namely:-

a. Glove Based
b. Vision Based
c. Low level features based

Model based approaches (Glove Based): Many models have already been applied to analysis, model and represent the hand shape, which gives us a description and make a wide range of human hand to be represented, and a large database for storing the extracted shape characteristics is needed as well[20]. Here in our work we have used glove based approach

Stage 2: It is further classified into four steps they are as mentioned:-

(i) Input test image: In this step input test image is taken in jpg format. The JPG file format, short for Joint Photographic Experts Group, is a type of image compression that works best with photographs and complex images. JPGs use a compression method that removes non-human-visibile colors from images to decrease file sizes. Be careful, though. If you decrease the quality of a JPG too much, you will begin to lose important color information that cannot be recovered. The JPG file format also allows you to save progressive JPGs, which will load in stages.

(ii) Pre-Processing and segmentation of image: In this step RGB image is now converted to gray scale image. RGB images are normal coloured images and grey scale images are those images which have various shades of grey colour in between black and white The various colour models that can be used for segmentation purpose are RGB, HSV (hue, saturation, and value), and CIE-Lab.

In our Hand Gesture Recognition system we will be using this second method for modeling our system. Here we have taken 10 snaps of every characters of American Sign Language.. There are different types of models available for Pattern Recognition [1].

- Statistical or decision theoretic or discriminate method.
- Syntactic or Grammatical or structural approach.
- Templates
- Features

Database Description

For our database, we have taken hand gesture photos using American Sign Language. The database consists of 24 characters of sign language sign. The whole system will work online i.e. we will feed test image as an input and the system will recognize which image character has been fed as input. Letter j and z are discarded because that can be described dynamically only. Our whole system works on static data and has dependency on data.

We have taken 340 images, 10 per gesture. Images that are captured need to have uniform dark colour background that can be black with a white colour rubber glove on hand as in contrast. We had done so that we can minimize noise and unwanted data and the segmentation process can be done easily. To have better result, user must wear a black colour cloth around arm till wrist from the shoulder to match with the black background and segmentation can be achieved easily. Covered arm and the background should be of similar colour.

Images of the hand gesture are captured using any camera. colour background with the black colour.

VII. Results & Analysis

The proposed algorithm is applied on 340 images with 34 different gestures which include all the static alphabets as per ASL and also on alphabets (0-9). With the help of defined features and encoded sequence, we can successfully recognize the 34 different hand gesture patterns.Furthermore these recognized gestures can be used for human computer interaction.

The graphical user interface of the final output, in which the test image is compared with the reference image. The experimental data is shown,starting with the input hand gesture image, along with the number of peaks in the graph. Under single orientation, either vertical or horizontal, the graph is unique.

The table shows the result of 340 images tested through this algorithm. Out of 340 images, it has correctly identified 330 images and falsely identified the remaining 10 cases, gives the success rate of 97.058% approximately with average computation time of 2 second for recognizing single image in image sequence. The algorithm is based on simple shape based feature calculation which provides us with the comfort of implementation.

The result is best suited for system application execution and is one of the best effort made for human machine interface.Suppose if we give D as the input then it matches the pattern with the stored optimized database,once the pattern matching is done it generates the graph of the input image and as per the instruction set in the program the system execution is done,here in case of D the desktop is executed.Similarly if we give A then avast antivirus is executed.The mentioned below table shows the system application execution:

Table 4.1: Directives to the system Executed as per defined and Recognized Gesture

INPUT SYMBOL	**Gesture Recognized**	**System Application Executed**
A	A	ANTIVIRUS (avast)
B	B	PICASA
C	C	CONTROL PANEL
D	D	DESKTOP
E	E	EXCEL
F	F	FOLDER CREATE
G	G	GOOGLE CHROME
H	H	HIBERNATE
I	I	INTERNET EXPLORER
K	K	POWER POINT
L	L	LOCK
M	M	MEDIA PLAYER
N	N	NOTEPAD
O	O	PAINT
P	P	PDF
Q	Q	PICTURE MANAGER
R	R	RESTART
S	S	SHUT DOWN
T	T	TORCH
U	U	µ torrent
V	V	VLC
W	W	WORD FILE
X	X	X to y converter
Y	Y	IDM
0	0	Library
1	1	Onenote
2	2	Cc cleaner
3	3	Downloads
4	4	Wondershare
5	5	M Blaze
6	6	Calculator
7	7	Game
8	8	IDM
9	9	My document

INPUT SYMBOL	No. of Input Samples	System Application Execution(percentage)
		Recognition Rate
A	10	100%
B	10	100%
C	10	100%
D	10	80%
E	10	100%
F	10	100%
G	10	100%
H	10	100%
I	10	100%
K	10	100%
L	10	100%
M	10	100%
N	10	100%
O	10	100%
P	10	100%
Q	10	80%

R	10	100%
S	10	100%
T	10	100%
U	10	80%
V	10	100%
W	10	100%
X	10	100%
Y	10	100%
0	10	100%
1	10	60%
2	10	100%
3	10	100%
4	10	100%
5	10	100%
6	10	100%
7	10	100%
8	10	100%
9	10	100%

The result of the output are as follows:-

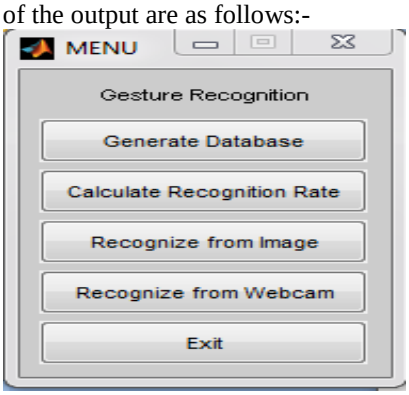


Fig.4.2 Menu

The above dialogue box states five different thing:-

(i) The first dialogue box Generate Database is used for loading and training the Data

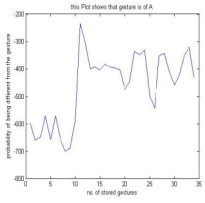
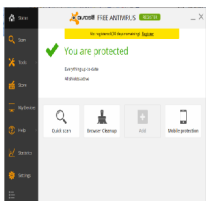
(ii) The second dialogue box Calculate Recognition Rate is used to calculate the percentage.Here out of 340 images 330 images obeys the pattern matching and 10 images were not recognised.In this approach 97.058% image is recognised.

(iii) The third dialogue box Recognized from image is used to give individual input in jpg format which is used for pattern patching from the stored optimized image.

(iv) The fourth dialogue box Recognize from webcam is used to capture image,here I have used youcam to capture images and that captured image is taken as input here and used for pattern matching from the stored optimized image.

(v) The fifth is EXIT as the name suggest is used to come out of the program.

Sample Output

VIII. Conclusion

Gesture recognition has been recent advancement which has been used by many big IT companies to develop applications for the same also. Gestures have been used to define execution of system application by big companies like IBM, Microsoft etc.

Most of them have used random gestures to do so and has been successful too. But they have not used some basic methodology of defining Gestures and also their Gestures are dynamic too. In this work an attempt has been made to use a predefined format of human Gestures named as American Sign Language.

As per the American Sign Language (ASL) and database has been created and collected at one place. Here, Glove based approach has been used. First of all the gestures has been captured using any normal camera wearing a good quality glove with black background which is given as the input database. We have taken total of 340 images i.e. 10 samples of all 34 Gestures which include 24 Alphabet Gestures and 10 Numerical Gestures. Two dynamic Gestures J & Z has been omitted because of their randomly changing patterns as per ASL.

Then neural network training of the database is done using Hidden Markov model (HMM). HMM helps to find the probability of similarity in different images stored in the database. Then the database has been optimized means out of ten given input the best possible five image is taken which is used for pattern matching and their size is also reduced.

For pattern matching, input Gesture is taken in jpg format which is converted into gray scale image, which is further optimized by converting it into PGM format which is used for pattern matching with the stored optimized image.

For all the input given graphs has been plotted which prove that the matched gestures has least log sequence. Based on the input and the gesture recognized then the system is given command to execute the pre specified instruction.

In this work we have taken 340 images out of which 330 images were correctly recognized and 10 images were falsely recognized. The total recognition rate is 97.058%.

Future Work

Gesture recognition has been growing and this effort is not an end, the work done and proposed in this dissertation can also be extended using many other techniques like:

1. The input gestures are static using still images , in future dynamic or moving images can also be taken so that run time issues can also be addressed
2. We have taken only JPEG & PNG as input; the same work can also be done using TIFF images.
3. HMM has been used for pattern matching, there are many techniques which can also be applied in future to get better results.
4. Better system application or Input Output Devices can also be controlled using the in future.

5. Using better noise removal techniques the results may also be further enhanced.

REFERENCES

[1] Henrik Birk and Thomas Baltzer Moeslund, "Recognizing Gestures From theHand Alphabet Using Principal Component Analysis", Master's Thesis,Laboratory of Image Analysis, Aalborg University, Denmark, 1996.

[2] Andrew Wilson and Aaron Bobick, "Learning visual behavior for gesture analysis", In Proceedings of the IEEE Symposium on Computer Vision, Coral Gables, Florida, pp. 19-21, November 1995.

[3] V. Pavlovic, R. Sharma, and T. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review", IEEE Trans. Pattern Analysis and Machine Intelligence, 19(7):677–695, July 1997.

[4] J. Lee and T. L. Kunii, "Model-based analysis of hand posture", IEEE Computer Graphics and Applications, pp. 77-86, September 1995

[5] J. J. Kuch and T. S. Huang, "Vision based hand modelling and tracking", In Proceedings of International Conference on Computer Vision, (Cambridge,MA), June 1995

[6] J. Lee and T. L. Kunii, "Constraint-based hand animation", In Models and techniques in computer animation, pp. 110-127, Tokyo: Springer-Verlag, 1993

[7] L. R. Rabiner , "A tutorial on hidden Markov models and selected applications in speech recognition", In Proceedings of the IEEE, 77(2), 257–286,1989.

[8] A. Wilson and A. Bobick, " Parametric hidden markov models for gesture recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence,21(9), 884–900, 1999.

[9] V. Pavlovic, R. Sharma and T.S. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review", IEEE Transaction on Pattern Analysis and Machine Intelligence, 19(7), pp 677–695,1997.