International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:3, 2007

Pakistan Sign Language Recognition Using Statistical Template Matching

Aleem Khalid Alvi, M. Yousuf Bin Azhar, Mehmood Usman, Suleman Mumtaz, Sameer Rafiq, Razi Ur Rehman, Israr Ahmed

***Abstract*—** Sign language recognition has been a topic of research since the first data glove was developed. Many researchers have attempted to recognize sign language through various techniques. However none of them have ventured into the area of Pakistan Sign Language (PSL). The Boltay Haath project aims at recognizing PSL gestures using Statistical Template Matching. The primary input device is the DataGlove5 developed by 5DT. Alternative approaches use camera-based recognition which, being sensitive to environmental changes are not always a good choice. This paper explains the use of Statistical Template Matching for gesture recognition in Boltay Haath. The system recognizes one handed alphabet signs from PSL.

***Keywords*—**Gesture Recognition, Pakistan Sign Language, Data Glove, Human Computer Interaction, Template Matching, Boltay Haath

I. Introduction

THIS system is a computerized sign language recognition system for the vocally disabled (deaf and dumb) who use sign language for communication. The basic concept involves the use of special gloves connected to a computer while a disabled person (who is wearing the gloves) makes the signs. The computer analyzes these gestures and synthesizes the sound for the corresponding word or letter for normal people to understand.

Since only single handed gestures have been considered in this project it is obviously necessary to select a *subset* of PSL to be considered for implementation of Boltay Haath (Boltay Haath is an Urdu phrase meaning 'Talking Hands') as it would take vast amounts of time to sample most or all of the signs in PSL [1].

Data gloves are special gloves equipped with sensors for detecting finger bend, hand position and orientation. They were conceived to allow a more natural interface to computers. However, the extension of their use for recognizing sign language is possible [2]. But progress in the recognition of sign language, as a whole has been limited [3].

The recognition engine is mainly based on three algorithms: dynamic pattern matching, statistical classification, and neural networks (NN). [4 - 6]

Traditionally, the technology of gesture recognition was divided into two categories, vision-based and glove-based methods. In vision-based methods, computer camera is the input device for observing the information of hands for fingers. However, the computation complexity in tracking of hands has several bottlenecks, such as feature extraction, objects need separation from background, fingers motion tracking, etc. Thus, it is difficult to achieve real time operation; we have turned to glove-based technique which is more practical in gesture recognition [7].

The benefits of sign language understanding systems are often debated and not made clear. A functioning system would provide an opportunity for the deaf to communicate with non signing people without the need for an interpreter. Although it is argued that a keyboard connected to the speech synthesizer could be used for this purpose, it is not the natural interface for signer and places an intermediary into the dialogue [8].

II. Components of the System

The basic components of the Boltay Haath system are given below:

A. *Modules for Gesture Input* – Get state of hand (position of fingers, orientation of hand) from glove and convey to the main software.

B. *Gesture Preprocessing Module* – Convert raw input into a process-able format for use in pattern matching. In this case, scaled integer values ranging from 0 to 255.

C. *Gesture Recognition Engine* – Examines the input gestures for match with a known gesture in the gesture database.

D. *Gesture Database* - Contains the necessary information required for pattern matching as well as a gesture-to-text dictionary.

Manuscript received November 20, 2004. This work was undertaken in the Sir Syed University of Engineering & Technology, Karachi, Pakistan.

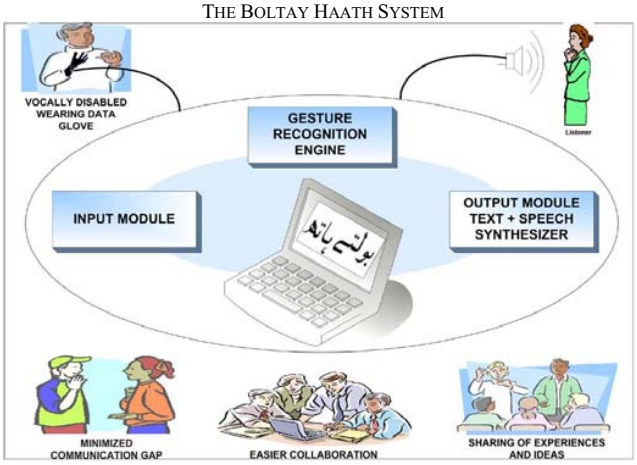
Aleem Khalid Alvi is an Assistant Professor in the Sir Syed University of Engineering & Technology, Karachi, Pakistan (phone: 92-021-4982106; fax: 92-021-4982393; e-mail: akalvi@ssuet.edu.pk).

M. Yousuf Bin Azhar., Mehmood Usman, Suleman Mumtaz, Sameer Rafiq, Razi Ur Rehman and Israr Ahmed are undergraduate students at the Sir Syed University of Engineering & Technology, Karachi, Pakistan (www.boltayhaath.cjb.net).

E. *Speech Synthesis Module* – Converts word / letters from moving alphabets for English as well as Urdu since PSL obtained after gesture analysis into corresponding sound

The following diagram best describes the top level components and benefits of Boltay Haath.

FIGURE I
THE BOLTAY HAATH SYSTEM



III. THE MODEL

The statistical model used in Boltay Haath is the simplest approach to recognize postures [9] (static gestures). The model used is known as "Template Matching" or "Prototype Matching". The idea is to demarcate different gestures by calculating the mean (μ) and standard deviations (σ) of all the sensors for a gesture and then those input samples that are within limits bounded by an integral multiple of standard deviation are recognized to be correct. Gesture boundary [10] for each sensor is defined as,

$$\mu \pm k\sigma \ , \ k = 1,2,3,\ldots \tag{1}$$

$$\mu_{(l,m)} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2}$$

$$\sigma_{(l,m)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \tag{3}$$

Here, x_i is the i^{th} sample, k is the integral multiple of σ, n is the number of samples, $\mu_{(l,m)}$ is the mean of the l^{th} sensor of the m^{th} gesture and $\sigma_{(l,m)}$ is the standard deviation of the l^{th} sensor of the m^{th} gesture.

IV. TRAINING

The system was trained by training data obtained from 6 different signers. Initially training data was collected for the alphabets for English as well as Urdu since PSL contains both types of signs. This was done due to the limitations of the input device i.e., the DataGlove about abduction status and the absence of any kind of input about the location of the glove in space. Hence a training set of more than 2500 samples was collected.

A training sample consists of five values ranging from 0 to 255 each representing the state of the sensor on all five fingers of the glove. The sensors for roll and pitch have been ignored since their values do not uniquely identify an alphabet sign.

This training data was then processed i.e., mean (μ) and standard deviation (σ) was calculated for all five sensors of each gesture in the training set. The resultant μ,σ pairs were stored in the gesture database for later use in gesture recognition.

V. GESTURE RECOGNITION ENGINE

After training, test samples are provided to the Gesture Recognition Engine which analyses them using the statistical model described previously. The upper and lower limits for the value of a sensor for a particular gesture are defined using the standard deviation for that sensor previously calculated. For enhancing the accuracy of gesture recognition, various integral multiples of σ are used, denoted by k in (1). The limits for any given gesture are defined as:

$$\text{Upper Limit}_{(l,m)} = \mu_{(l,m)} + k\sigma_{(l,m)} \tag{4}$$

$$\text{Lower Limit}_{(l,m)} = \mu_{(l,m)} - k\sigma_{(l,m)} \tag{5}$$

Given the above mentioned criteria any given input can be classified as a particular gesture if all the sensor values of the test sample lie within these limits. These values are retrieved from the gesture database.

The values of k used for gesture recognition in Boltay Haath range from 1 to 3, providing tolerances ranging from 2σ to 6σ. The performance achieved by varying the values of k is discussed later in this paper.

Sometimes due to ambiguity between two gestures the system may produce two outputs. To cater to this problem the method of Least Mean Squares (LMS) is used.

VI. LMS FOR REMOVING AMBIUITY

There are cases where more than one gestures are candidates for output. To overcome this type of situation the system calculates Least Mean Squares (LMS) [11] of all the candidate gestures and then selects the one with minimum LMS value [12]. The use of LMS is justified by the results. Analyzing the performance of the system it has been observed that the use of LMS provides 60 % accurate results.

LMS value is calculated as,

$$\text{LMS} = \sum (x_i - \mu_i)^2 \tag{6}$$

Here, x_i denotes the sensor value of the i^{th} sensor from test sample, μ_I denotes mean value for the i^{th} sensor. LMS for each candidate gesture is calculated and the gesture with lowest LMS value is selected as the output.

VII. Implementation

Boltay Haath system has been developed in C# using Visual Studio .Net 2002. The gesture database was maintained on a MS Access database file. All the processing on the data [13] was done using SQL queries. The results were verified in real time. Windows being the platform for the project, all the user interface and input components were standard windows objects [14]. Microsoft Speech SDK 5.1 was used for speech output. However the phonemes had to be modified in order to produce sound matching the accent and pronunciation of the people of Pakistan.

The complete working of the project involves the use of a DataGlove5 connected to a computer [15] and software modules for preprocessing [16]. It performs analysis of data to minimize the variations, analysis of gestures, extraction of words / letters from database and generation of the corresponding sounds.

VIII. Performance

The performance of Boltay Haath was evaluated by providing various test cases [17] for both English and Urdu gestures. Using various values of k in (1) the accuracy of the system was determined. The system was also evaluated with and without the use of LMS to handle ambiguities among similar gestures. The results obtained are presented in Tables 1 and 2. Tolerance values range from 2 to 6 σ depending on the value of k.

PSL contains some signs which are either too similar to other signs making them difficult to distinguish or contain aspects which cannot be read by the DataGlove without additional sensors. By labeling these signs as 'ambiguous' and excluding them from the results it has been observed that the actual accuracy is far higher than the observed accuracy. So if appropriate sensors are added to the system, performance will increase considerably.

TABLE I
Performance Results (English Alphabets)

Criteria		Accuracy (%)		
	Tolerance	2 σ	4 σ	6 σ
Including ambiguous signs	With LMS	23.8	68.5	71.3
	Without LMS	21.3	36.2	15.4
Excluding ambiguous signs	With LMS	25.4	73.3	78.2
	Without LMS	23.4	43	20

In Table I, it can be observed that the best overall performance achieved is 71.3 % when 6σ is used. However, when ambiguous gestures are ignored, the accuracy increases to 78.2 %. In both cases, the results turn out to be poor if LMS is not used.

TABLE II
Performance Results (Urdu Alphabets)

Criteria		Accuracy (%)		
	Tolerance	2 σ	4 σ	6σ
Including ambiguous signs	With LMS	25.9	67.8	69.1
	Without LMS	21.7	26.8	15.1
Excluding ambiguous signs	With LMS	31.1	81.4	85
	Without LMS	28.8	33.5	21.5

In Table II, it can be observed that the best overall performance achieved is 69.1 % when 6σ is used. However, when ambiguous gestures are ignored, the accuracy increases to 85 %. In both cases, the results turn out to be poor if LMS is not used.

Chart I
Trend In Accuracy Change

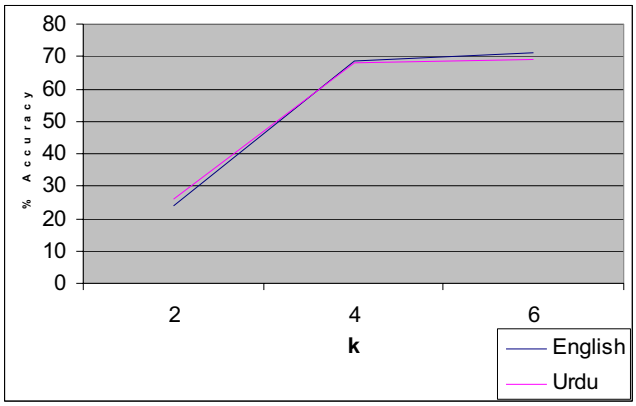


Chart I shows the trend in accuracy when k of (1) is increased. The accuracy of the system increases drastically when the limit is changed from 2σ to 4σ . After that the performance improves but not very rapidly and stabilizes somewhat at 6σ.

Chart II
Test Results for Urdu Signs

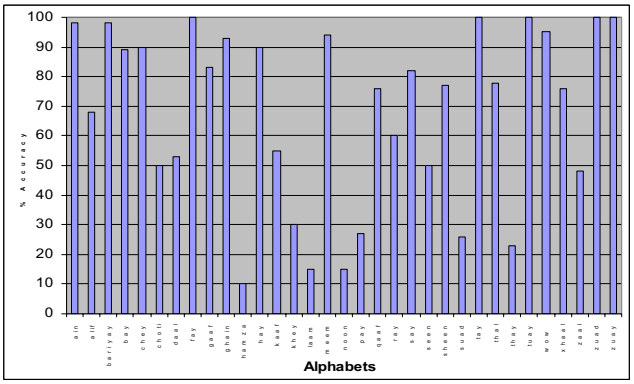


Chart II shows the test results for Urdu signs. The test cases did not include dynamic or moving gestures.

CHART III
TEST RESULTS FOR ENGLISH SIGNS

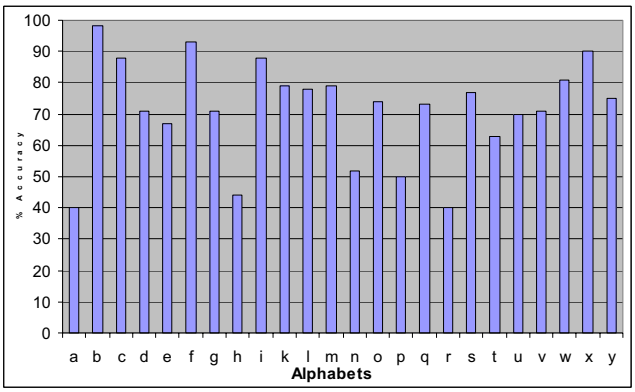


Chart III shows the number of correctly recognized gestures against all test cases of the particular alphabet. As can be observed the letters 'H' and 'R' give poor accuracies. This is because the letter 'R' and 'H' are very similar as can be seen in the hand-shapes below. Since the Boltay Haath system does not understand the abduction between fingers, it is very difficult to distinguish between such gestures.

FIGURE II
AMBIGUOUS SIGNS

The following graph (Graph 2) shows the variation in recognition accuracy among the various alphabets.

IX. CONCLUSION

Deaf and Dumb people rely on sign language interpreters for communication. However, they cannot depend on interpreters every day in life mainly due to the high costs and the difficulty in finding and scheduling qualified interpreters. This system will help disabled persons in improving their quality of life significantly.

The automatic recognition of sign language is an attractive prospect; the technology exists to make it possible, while the potential applications are exciting and worthwhile. To date the research emphasis has been on the capture and classification of the gestures of sign language and progress in that work is reported. This project will be a valuable addition to the ongoing research in the field of Human Computer Interface (HCI).

ACKNOWLEDGMENTS

We would like to thank Mr. Waleed Kadous of University of New South Wales, Australia for his support and guidance regarding the various aspects of gesture recognition, neural networks and statistical template matching.

Further we would like to thank Mr. Richard Geary and Mr. Ali Akber of Deaf Reach, Pakistan for teaching us Pakistan Sign Language, its usage, vocabulary and basic signs.

Also we would like to acknowledge the support of Mr. Israr Umer of the Special Education Dept., University of Karachi, Pakistan for his information on two PSL signs and related material.

REFERENCES

[1] Dr. Nasir Sulman, Sadaf Zuberi, "Pakistan Sign Language – A Synopsis", Pakistan., June 2000.
[2] I. Wachsmuth, T. Sowa (Eds.), "Towards an Automatic Sign Language Recognition System Using Subunits", London, April 2001, pp. 1-2
[3] Waleed Kadous, "GRASP: Recognition of Australian sign language using Instrumented gloves", http://www.cse.unsw.edu.au/~waleed/thesis/thesis.html, Australia, October 1995,pp. 1-2.
[4] Andrea Corradini, Horst-Michael Gross. 2000, "Camera-based Gesture Recognition for Robot Control",2000 IEEE 133-138.
[5] Andrea Corradini ,Horst-Michael Gross. 2000, "A Hybrid Stochastic-Connectionist Architecture for Gesture Reognition",2000 IEEE 336-341.
[6] Vesa-Matti Mantyla, Jani Mantyjarvi, Tapio Seppanen, Esa tuulari. 2000, "Hand Gesture Recognition of a mobile device user",2000 IEEE 281-284.
[7] Sim Oni, "Gesture Recognition Using Neural Network", Taiwan, 2000, pp. 1-2.
[8] Richard Watson, "A survey of Gesture Recognition Techniques Technical Report", Trinity College, Dublin, July 1993, pp. 6
[9] The Webopedia Website, www.webopaedia.com
[10] Aleksander, I. and Morton, H., An Introduction to Neural Computing, Chapman & Hall, London, 1990. Amari, S. I., "Learning patterns and pattern sequences by self-organizing nets," IEEE Trans. Comput., vol. 21, pp. 1197-1206, 1972.
[11] Dictionary Dot Com, http://www.dictionary.com/
[12] Dictionary Dot Com reference, http://dictionary.reference.com/
[13] Barbara Liskov , Program development in java, pg. 356,chap 11
[14] Ian Sommerville, Software Engineering, pg. 8, chap 1
[15] The 5DT Website, www.5dt.com
[16] Dictionary Online, http://dictionary.reference.com
[17] S. Sidney Fels. Glove-TalkII: Mapping Hand Gestures to Speech Using Neural Networks -- An Approach to Building Adaptive Interfaces. PhD thesis, Computer Science Department, University of Toronto, 1994., pp 35-42
[18] Murakami and Taguchi, Gesture recognition using recurrent neural networks. In CHI '91 Conference Proceedings, pages 237--242. Human Interface Laboratory, Fujitsu Laboratories, ACM, 1991.
[19] Peter Vamplew. The SLARTI sign language recognition system: A progress report. University of tasmanis, Australia, Pp. 1-3
[20] Waleed Kadous, "GRASP: RECOGNITION OF AUSTRALIAN SIGN LANGUAGE USING INSTRUMENTED GLOVES ", Australia, OCTOBER 1995 , pp. 1-2, http://www.cse.unsw.edu.au/~waleed/thesis/thesis.html
[21] K.S. Fu. Syntactic Pattern Recognition, Prentice-Hall 1981, Pp 75-80
[22] K.S. Fu. and T.S. Yu. Statistical pattern Classification using Contextual Information, Recognition and image Processing Series, Research Studies Pres, 1980.
[23] David J, Sturman, Whole hand input, Ph.D. Thesis, MIT, 1992, Pp. 14-56
[24] Dean Rubine, Automatic Recognition of gestures, PhD, thesis, Carnegie Mellon University , December 1991, Pp. 90-156