Recognition of Two-handed Arabic Signs using the CyberGlove

Mohamed A. Mohandes
Electrical Engineering Department
King Fahd University of Petroleum and Minerals
Dhahran, 31261, Saudi Arabia
mohandes@kfupm.edu.sa

***Abstract*--Sign language maps letters, words, and expressions of a certain language to a set of hand gestures enabling an individual to communicate by using hands and gestures rather than by speaking. Systems capable of recognizing sign-language symbols can be used as a means of communication between hearing-impaired and vocal people. This paper represents the first attempt to recognize two-handed signs from the Unified Arabic Sign Language Dictionary using the CyberGlove and support vector machines. Principal Component Analysis is used for feature extraction. 20 samples of each of 100 two-handed signs were collected from an adult signer. 15 samples of each sign were used for training a Support Vector Machine to perform the recognition. The performance is obtained by testing the trained system on the remaining 5 samples of each sign. A recognition rate of 99.6% on the testing data was obtained. When more signs will be considered, the support vector machine algorithm must be parallelized so that signs are recognized on real time.**

Keywords—Arabic sign language; recognition; support vector machine; principle component analysis.

I. INTRODUCTION

Developing a pattern recognition system for sign language interpretation is a very difficult process. One difficulty is that use of traditional programming paradigms makes the system overwhelmingly complex and hence impractical. This dictates resorting to machine-learning methods. Another difficulty encountered is the interface issue. Ideally, the interface should deliver accurate measurements to the processing machine, have low cost, and provide input in a form that requires low pre-processing overhead. Building a system that satisfies these three requirements is very challenging. Hence, design compromises must be done to build a practical system.

Interfaces in sign language systems can be categorized as direct-device or vision-based. The direct-device approach uses measurement devices that are in direct contact with the hand such as instrumented gloves, flexion sensors, styli and position-tracking devices. On the other hand, the vision-based approach captures the movement of the singer's hand using a camera that is sometimes aided by making the signer wear a glove that has painted areas indicating the positions of the fingers or knuckles. The main advantage of vision-based systems is that the user isn't encumbered by any complex devices. Their main disadvantage, however, is that they require a large amount of computation just to extract the hands position before performing any analysis on the images. This paper deals only with the directed-devise methods.

The first widely known instrumented glove is the Digital-Data-Entry Glove [1, 2]. It was originally proposed as an alternative input device to the keyboard and worked by generating ASCII characters according to finger positions. The gloves had finger flex sensors, tactile sensors at their tips, orientation sensors and wrist-positioning sensors. The VPL-DataGlove used novel optical flex sensors that had fiber optic cables with a light at one end and a photodiode at another. A simplified version of the latter is called the Z-glove. It uses fiber optic devices to measure the angles of each of the first two knuckles of the fingers and is usually combined with a Polhemus tracking device. The Z-glove was the first commercially available instrumented glove. The Exon-Dextrous-Hand-Master was developed afterwards with 8 bits of accuracy, 20 degrees of freedom and a measurement frequency of 200 Hz [3]. The PowerGlove is a highly cost effective alternative to other instrumented gloves but less accurate [3]. It is based on VPL's glove and only measures the position in the three dimensional Cartesian space and the roll while other gloves measure the pitch and yaw as well.

Section II of this paper discusses briefly previous work related to sign language recognition. Section III introducers the proposed system, while Section IV discusses the preprocessing and feature extraction. Section V highlights the recognition of the Arabic sign language, and Section VI describes case study of the developed system. Section VII concludes the paper.

II. RELATED WORK

David L. Quam used a DataGlove Model 2 in addition to a Polhemus tracker [4]. Twenty two signs from the American Sign Language (ASL) were provided by two signers, one is the right handed male and the other is the left handed female. Most of the signs are letters or numbers in addition to two selected words. Due to the small set of signs, he used the finger flexions and hand orientation directly as features. This system was only able to identify static signs, thus had limited applications.

 ISBN: 978-1-61208-101-4

Fels and Hington developed a system called Glove-Talk [5]. They used a VPL-DataGlove, having 2 sensors per finger, and a Polhemus tracker. They used five neural networks to connect the gloves to a speech synthesizer. The Glove-Talk project vocabulary consists of 203 words. The Glove-Talk project gives an accuracy of 94%.

Waleed Kadous developed a system called "GRASP" for the recognition of the Australian Sign Language [6]. He used a PowerGlove for collecting data. He used the energy, time, bounding boxes and simple time division over a specified number of segments as features. 6,650 samples of signs were collected from 5 different people. With 95 different signs, the accuracy of the system was about 80%.

Waldron and Kim used a DataGlove with a Polhemus tracker to obtain the hand shape and position [7]. They developed a two-stage neural network system to recognize isolated ASL signs. The first stage recognizes the sign language phonology consisting of 36 hand shapes, 10 locations, 11 orientations, and 11 hand movements using four different neural networks. The second stage uses the recognized phonemes from the beginning, middle, and end of the sign as input to identify the actual sign. Six signers generated 14 signs. The overall performance of sign recognition was 86%.

Sagawa, Takeuchi, and Ohki developed Japanese Sign Language recognition system [8]. The sign is represented as a combination of basic components of gestures. These components are called cheremes. The system uses two CyberGloves and two trackers. A total of 14 cheremes are recognized by the system. The recognized cheremes are sent to the recognition part of sign language morpheme. The system is used to recognize 60 sings. Twenty samples were collected from each signs, 10 samples used for training and 10 for testing. The recogntion rate is 97.6%.

Kim, Jang and Bien investigated the recognition of the Korean Sing Language (KSL) [9]. Two DataGloves and two Polhemus trackers are used. KSL signs can be formed by combining a small number of basic gestures. 25 basic gestures are considered with 14 basic hand shapes. The hand shapes are recognized using a fuzzy min-max neural network. For the recognition process, the direction type is identified and then the hand shape of the motion is recognized. The recognition rate reaches 85% for this algorithm.

Jiangqin and co-investigators used a CyberGlove and a 3-D tracker to recognize signs from the Chinese sign language [10]. Each sign of the 3300 Chinese sign language is characterized by posture, orientation, position and motion trajectory. The number of postures for the right hand and the left hand are 14 and 7, respectively. They used multilayer perceptron to code the data as the input to the Hidden Markov Models. The recognition rate of the samples is over 90%.

Mohandes and Al-Buraiky used a PowerGlove to recognize single-handed Arabic signs [11]. For feature extraction, they used time division where the data is divided into segments and the average of each segment is calculated for each sensor. SVM is used for the recognition process. 36 samples of each of 120 signs were collected from a deaf signer, 18 used for training and the remaining 18 are used for testing. A recognition rate of about 70% was achieved. The CyberGlove was used in our lab for the recognition of single-handed Arabic Signs [12], while this paper uses two CyberGloves to recognize two-handed signs from the Arabic sign language using two CybeGloves and support vector machine. The SVM is optimal on the sense that it maximizes the separation margins among classes and therefore, it is expected to outperform other methods on the recognition of all sign languages.

III. THE PROPOSED SYSTEM

The proposed system consists of two CyberGloves, two hand tracking systems, and data acquisition software. The CyberGlove is a fully instrumented glove that provides 22 high-accuracy joint-angle measurements as shown in Figure 1. It uses proprietary resistive bend-sensing technology to accurately transform finger motions into real-time digital joint-angle data. Each sensor is extremely thin and flexible being virtually undetectable in a lightweight elastic glove. The CyberGlove has been used in a wide variety of real-world applications, including digital prototype evaluation, virtual reality biomechanics and animation. In addition to the CyberGloves two hand tracking devices are added to measure the location (x, y, z) and orientation (yaw, pitch, roll) of each hand with reference to a fixed point. There are three main components of a tracking device: a transmitter generating a signal, a sensor that receives the signal, and a control box for signal processing and connection to the computer.

The tracking device used in this paper is the Flock of Bird (FOB). It is used to track the position and orientation of up to thirty sensors simultaneously by a transmitter. Each sensor is capable of making from 20 to 144 measurements per second of its position and orientation when it is located within 4 feet of its transmitter. The sensor is fixed at the wrist of the CyberGlove, as shown in Figure 1.

Each CyberGlove provides 22 sensor signals and 6 signals are provided from each FOB, thus a total of 56 measurements are provided from the two gloves and the two hand tracking devices while the signer is performing the signs. These measurements are sent through the serial ports and stored in a human readable format (ASCII text).

Figure 1. The CyberGlove

IV. PREPROCESSING AND FEATURE EXTRACTION

A continuous stream of frames is generated by the gloves and the hand trackers as the hands perform the sign. Each frame bears an instantaneous measurement of each of the 56 signals for the two hands. Signs have different lengths. Even samples of the same sign performed by the same signer may have different lengths as well. The classification system requires the same number of inputs. Therefore, time division is used to make all signs have the same number of components. In this paper, the duration of the sign is divided into 10 segments. The mean and standard deviation are calculated from each signal in each of the 10 segments. Thus, each signal is represented by 20 values of the means and standard deviations of the segments. Thus, the signals from the 56 sensors will be represented by 1120 values. Principal Component Analysis (PCA) is used to reduce the dimensionality of the data. Thus PCA is used to provide features for the classification machine.

Given a d-dimensional vector representing the mean and standard deviation from each segment of the raw data, the Principal component analysis (PCA) can be used to find a subspace whose basic vectors correspond to the maximum-variance direction in the original space [13]. Let W represent the linear transformation that maps the original d-dimensional space onto a lower dimensional space F-dimensional feature subspace. The new feature vectors $y_i \in R^F$ are defined by $y_i = W^T x_i, \quad i = 1,\ldots,N$. the columns of W are the eigenvectors e_i obtained by solving the equation $\lambda_i e_i = Q e_i$, where $Q = XX^T$ is the covariance matrix, and λ_i is the eigenvalue associated with eigenvector e_i. Before obtaining the eigenvectors of Q, the vectors are normalized and the mean is subtracted from all normalized vectors. In this paper a different number of eigenvlaues have been used.

V. RECOGNITION OF ARABIC SIGNS

After feature extraction, the signs are ready for the classification. In this paper, Support Vector Machine (SVM) is used for the recognition of the signs. In this section SVM is briefly introduced. In its simplest form, SVM is based on constructing a hyperplane that separates two linearly separable classes with the maximum possible margin [14]. Assuming that there is a set of vectors x each having a label y and that the separating hyper plane is $w.x + b = 0$, where w is a weight vector and b is a constant. It can be shown that the decision function (hypothesis), which is based on the optimal hyperplane, takes the form:

$$f(x) = \sum_{i=1}^{l} \alpha_i y_i x_i * x + b \tag{1}$$

providing that the coefficients a_i maximize the following function:

$$L_D = \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i,j}^{l} y_i y_j \alpha_i \alpha_j x_i * x_j \tag{2}$$

where L_D is called the dual Lagrangian function and is obtained by deriving the dual of the optimization problem formulated to maximize the classification margin. A support vector machine for separating non-linearly separable data can be built by first using a non-linear mapping that transforms data from the input space to a higher space called the feature space, and then using a linear machine to separates them in the feature space. Mapping to the feature space can be performed by replacing the dot products with a kernel function $K(x,z) = \phi(x).\phi(z)$ (where $\phi : X \to F$ is a non-linear mapping from input space X to feature space F). A number of different kernel functions can be found in the literature. The kernel function used in this paper is the Radial Basis Kernel defined as

$$K(x,y) = e - \gamma \|x - y\|^2 \tag{3}$$

By replacing the dot product with the kernel function the decision function (the hypothesis) becomes:

$$f(x) = \sum_{i=1}^{l} \alpha_i y_i K(x_i.x) + b \tag{4}$$

This allows the SVM algorithm to solve real-world problems where data can be non-linearly separable. Additionally, in feature space, some slackness can be introduced in the support vector machine developed above so that some error is tolerated. This is done by adding slack variables (representing violations of the margin constraints) to the cost function. With this modification the optimization problem becomes:

Minimize $w.x + C\sum_{i=1}^{l}\zeta_i^2$ (5)

subject to $y_i((w_i.x_i)+b) \geq 1-\varsigma_i, i=1,2,..l$ (6)

where $\varsigma_i, i=1,2,..l$ are slack variables and C is a penalty error constant coefficient whose best value is determined in practice by trail and error adjustment.

Solving for support vectors (optimizing the dual function) is a quadratic convex programming problem, for which many numerical solution algorithms exist. Figure 2 shows the structure of the SVM classifier.

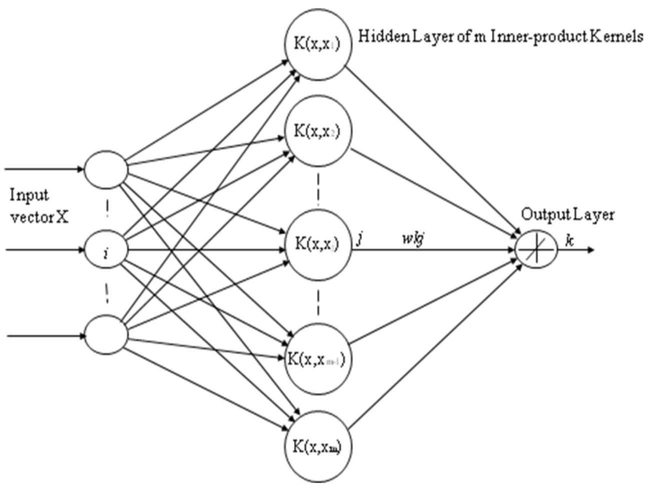

Figure 2. The Support Vector Machine Structure

The above formulations can be extended to the classification of n classes by one of the following methods :

1. Build *n* classifiers, each capable of separating patterns belonging to one class from all other patterns.
2. Build the *n*-class classifier by feeding input to each of the two-class classifiers and choosing the class corresponding to the maximum $f_k(x), k=1,2,....,n$

The multi-class problem can be solved in a direct manner as well as by generalizing the procedure used for the two-class case [14].

VI. CASE STUDY

A volunteer from the deaf community performed the signs to generate samples for the learning machine. The signer was chosen among adults to insure his fluency in sign language and the accuracy of the signs. The Signer performed twenty samples of each of 100 two-handed signs selected from the Arabic Sign Language Dictionary. The signer starts with his two hands resting on his side as shown in Figure 4. A button is pressed to signal the start of the sign. As soon as the sign is completed, the button is presses again. This process is repeated 20 times for each sig to produce a total of 2000 samples.

The 20 samples of each sign are divided into two parts: training and testing. The training data for each sign consists of 15 samples, and the remaining 5 samples are used for testing. The duration of every sign is different, even the samples of the same sign implemented by the same signer will take different time. Thus the number of data points from each sensor is different. For example, the 20 samples of the first sign have a number of data points that ranges between 15 and 22, as shown in Figure 3. The number of data points for the other signs ranges between a maximum of 40 and a minimum of 10. For the recognition machine, the number of data points on all samples of the signs should be the same. Therefore, a pre- processing step is added to unify the number of data points. The duration of each sign implementation is divided into 10 segments. If the number of data points is not a multiple of 10 then the extra points are distributed among the first segments. For example if a sign has 23 data points, then each segment will have 2 points except the first three segments will have 3 data points. The mean and standard deviation of each segment is calculated. Thus, each sensor signal is represented by 20 values which are the mean and standard deviation of the 10 segments of the signal.

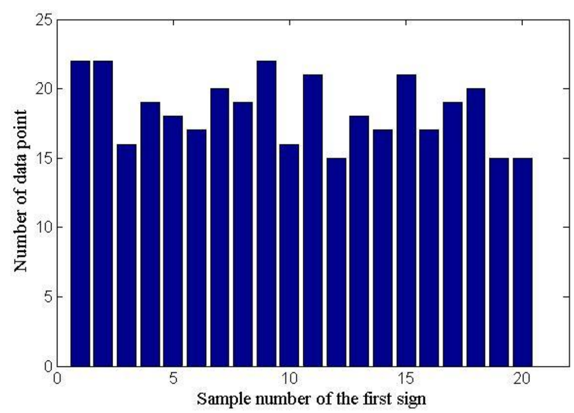

Figure 3. The number of data points of the 20 samples of the first sign

There are 56 measurements of the signs including 22 from each glove and 6 from each hand tracker. Each sensor signal is represented by 20 values which are the mean and standard deviation of each segment. Therefore, each sign is represented totally by a vector of length 1120 values. PCA is used for dimensionality reduction as explain in Section IV. Thus, the 2000 vectors representing 20 samples of each of 100 signs are normalized and the mean is subtracted. The covariance matrix is found and its eigenvectors and eigenvalues are calculated. The vectors, each of length 1120 values, which represent the samples of the signs, are transformed to a lower dimension by a linear transformation matrix formed by the eigenvectors of the covariance matrix. The number of eigenvectors chosen determines the size of the feature vector. The eigenvalues represent the variance of the data when projected onto the corresponding eigenvectors. Therefore, only the eigenvectors corresponding to the highest eigenvalues need to be considered. Figure 4. Shows the values of the first 100 eigenvalues, where the first eignevalue is 8,5629 and the 100th eigenvalue is 0.9133.

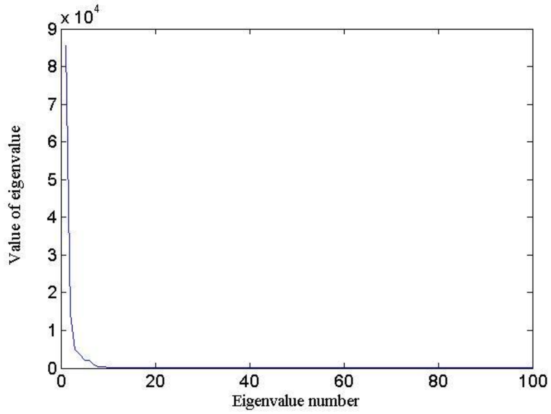

Figure 4. The eigenvalue spectrum

To determine suitable dimensionality of the feature vector for this application, an experiment is performed where several numbers of eigenvectors between 1 and 1120 are considered. The accuracy of the recognition system on the testing samples is calculated. Figure 5. shows the relation between the size of the feature vector and the accuracy of the recognition system on testing data. The figure indicates that a feature vector of size 70 would give the best performance. The figure indicates also that increasing the size of feature vector beyond 500 elements degrades the performance is it adds irrelevant information.

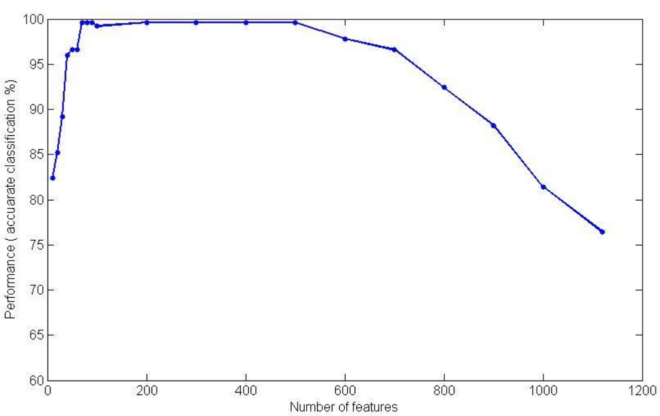

Figure 5. Performance with different number of feature vector

A support vector machine is trained using 15 samples of each of the 100 collected signs. The SVM has 70 inputs and one output. The output unit takes a value between 1 and 100. The value of the output unit indicates the sign that the input vector is assigned to. The Kernel function used in the SVM is the Radial Basis Kernel. After several experiments, it was found that a suitable values of the user defined SVM parameters for this application are the error penalty constant, C= 110, and $\gamma = 0.15$. The trained SVM is tested using the remaining 5 samples of each sign that have not been used in training. The performance of the trained SVM on the testing data lead to 99.6% correct classification where only two samples of the total 500 are misclassified as other signs. The two samples belong to one sign that is misclassified to another sign where the two signs differ only on the location of the hands, while the all the sensors have almost the same values as seen in Figure 6. The result shows the viability of SVM for the recognition of Arabic sign language.

Figure 6. Frames of the two signs that has been misclassified

To further test the performance of the developed system, another signer provided about 300 samples from 15 signs. The samples were sued to test the already trained SVM. However, the recognition rate did not exceed 63%. The low recognition rate could be due to the fact that the Arabic Sign language is not fully standardized and the two signers are from two different areas of the Kingdom. However, when samples from the second signer are included in the training, the recognition rate reaches 93%. This result indicates the need to collect samples from more signers to fully test the possibility of signer independent system.

The recognition process of each sign at this stage takes about 3 seconds. However, when a large number of signs are considered, the SVM has to be parallelized so that the recognition is done in real time.

VII. CONCLUSION

This paper is a contribution to the area of Arabic Sign Language recognition, which had very limited research. Two CyberGloves and two trackers are used to collect the signs data. The gloves and trackers provide 56 signals. The durations of the signs are different; therefore, the collected data is pre-processed by dividing the duration of each sign into 10 segments and taking the mean and standard deviation of each segment. Thus each sign is represented by a vector of length 1120 components which represent the mean and the standard deviation of each segment of the sign. Principal component analysis is used for feature selection. A support vector machine is used for the recognition. 20 samples of each of 100 different signs are collected from an adult deaf signer. 15 samples are used for training and 5 for testing. The PCA is used to get effective features from these signals. A feature vector of size 70 is shown to classify the signs very well. A recognition rate of 99.6% was achieved with only 2 signs are misclassified among the 500 signs used for testing which indicates a good performance of the developed system. For future work all signs of the Arabic Sign Language Dictionary will be recognized and several signers will provide the samples so that we reach a signer independent recognition system.

ACKNOWLEDGEMENTS

The work in this paper is partially funded by King Fahd University of Petroleum and Minerals Grant No. IN 100045 and by Prince Salman Center for Disability Research Grant No. 0065, and by the Arab Science and Technology Foundation Grant No. ALJ- FFC07. The author thanks M. Deriche for valuable discussions.

REFERENCES

[1] J. Kramer and L. Leifer, “The Talking Glove for non-Verbal Deaf Individuals”. Technical Report CDR TR 1990 0312, Center for Design Research, Stanford University, 1990.

[2] Peter Vamplew, “The SLARTI Sign Language Recognition System” University of Tasmania.

[3] D. J. Sturman and D. Zeltzer. “A Survey of Glove-Based Input”. IEEE Computer Graphics and Applications, 14(1), pp. 30-39, January 1994.

[4] D. Qaum, ”Gesture Recognition with a DataGlove,” IEEE Proc. Aerospace and Electronics Conference, vol.2, pp. 755-760, May 1990.

[5] S. Fels and G. Hington, "Glove-Talk: A Neural Network Interface Between a Data-Glove and a Speech Synthesizer," IEEE Trans. Neural Networks, vol. 4, pp. 2-8, Jan. 1993.

[6] W. Kadous, “*GRASP:* Recognition of Australian Sign Language Using Instrumented Gloves”. Bachelor’s Thesis, The University of New South Wales, 1995.

[7] M. Waldron, and S. Kim, “Isolated ASL Sign Language Recognition for Deaf Persons”, IEEE Trans. Rehabilitation Engineering, pp. 261-270, Sept. 1995.

[8] H. Sagawa, M. Takeuchi, and M. Ohki, “Description and Recognition Methods for Sign Language Based on Gesture Components,” Proc. IUI97, pp. 97-104, Orlando, Florida, ACM, 1997.

[9] J. Kim, W. Jang, and Z. Bien, “A Dynamic Gesture Recognition System for the Korean Sign Language (KSL),” IEEE Trans. System, Man and Cybernetics, pp. 354-359, vol. 26, no. 2, April, 1996.

[10] W. Jiangqin, G. Wen, S. Yibo, L. Wei and P. Bo, “A Simple Sign Language Recognition System Based On Data Glove,” Proc. ICSP ’98, pp. 1257-1269, 1998.

[11] M. Mohandes and S Al-Buraiky, “Automation of the Arabic Sign Language using the PowerGlove”, the ICGST International Journal on Artificial Intelligence and Machine Learning (AIML), V. 7, Issue 1, pp. 41-46, 2007.

[12] I. Al-Saihati, “Feature Extraction for Real Time Recognition of the Arabic Sign language”, MS thesis, KFUPM 2006.

[13] A. M. Martinez and A. C. Kak, “PCA versus LDA”, IEEE Transaction on Pattern Analysis and machine Intelligence, Vol. 23, No. 2, pp. 228-233, 2001.

[14] V. N. Vapnik, Statistical Learning Theory, John Wiley and Sons Inc., 1998.