Gesture recognition through HITEG data glove to provide a new way of communication

G. Saggio, P. Cavallo, A. Fabrizio, S. O. Ibe

Dept. of Electronic Engineering, Univ. of "Tor Vergata"
Via del Politecnico, 1 - Rome, Italy
saggio@uniroma2.it

Abstract—**Hands gestures recognition, by means of measuring apparatus, can provide a new way of human-computer interaction. Controlling different devices or speaking through a speech synthesizer can be time saving as well as an aid for impaired persons. In this work we performed the classification of 19 different gestures, evaluating three different methodologies: Support Vector Machines, Mahalanobis and Euclidean based classifiers.**

Keywords-gesture recognition; data glove; classification; support vector machine; gestures; euclidean distance classifier; mahalanobis distance classifier

I. INTRODUCTION

The technological development of recent years helps to solve some small problems on daily life, such as the simultaneous translate of word or phrases from a one language to another.

Until some time ago, in fact, we used the paper dictionary, which was neither very comfortable nor easy to use. Now there are some applications installed on portable devices that allow us to translate the majority of world languages in a simple and fast way.

However, today there is still a language closely linked to the dictionary: the Sign Language, which is the language adopted by the deaf, who communicate with hand gestures, body and facial expressions. Not being a spoken language, it is difficult to translate and understand for people who are not used to it.

Our idea concerns the realization of a device that would facilitate communication between deaf and hearing people, thanks to a system capable to recognize, in real-time, static hand gestures which encode the manual alphabet of sign language and reads, through a speech synthesizer, the letter corresponding to the recognized gesture. It can be considered a preliminary step toward the recognition even of dynamic gestures, for a proper "translation" from sign languages as the ASL (American Sign Language) or the LIS (Lingua dei Segni Italiana, i.e. the Italian Sign Language).

Besides that, our device can also be used as a controller to interact with a normal computer in order to speed up the interaction between the user and the software and to provide a more natural way of interaction.

II. STATE OF THE ART

In literature there are many works related to recognition of hand gestures, which differ in the tools used to capture the gesture, on how to classify the gesture and, most importantly, on the accuracy obtained. We can classify the instruments for gesture acquisition in two main different types.

The first is based on the acquisition of video signals by means of optical devices, such as webcam, capable of motion tracking of markers placed on each finger. After capturing the video, a computer vision's algorithm extrapolates the relevant data, in our case the configuration of the fingers of the hand and its position in the space [1]. Generally, the algorithms are very complex, have a high computational cost, and the cameras suffer from visual occlusion problems, so the recognition of the signs in real-time could be meaningfully affected.

The second gesture acquisition system uses the so called "data-glove", i.e. a supporting glove equipped with sensors, which measure the bending angles of each hand joints and the hand position in space.

After the gesture acquisition it follows the classification. Regarding this point, the most adopted classifiers have been Neural Networks, Hidden Markov Models and Support Vector Machines [2]. The first time that effort was made to recognize the movement of hands was in 1992. From that time on, there have been efforts to realize different systems of classification and Gesture recognition, like the work of Starner and Pentland [3]. They implemented a system to track hands and translate the motions into American Sign Language (ASL) in real time.

Takahashi et al. [4] used data glove for the first time in 1991 in order to recognize 46 symbols of Japanese Sign Language. It was the first time the approach of recognition was based on finding the principal component of a gesture (Principal Component Analysis) and the accuracy was estimated around 65%. Then in 1996 Jong-Sung used neural networks for the classification of static gestures with an accuracy of 85% over 25 symbols [5].

Recently in 2006, Xu Deyou realized a simulator in which using 15 hand postures of the Korean sign Language, it was possible to make actions on a machine, for example switching

on or off an engine or start the movement of a vehicle and so on. In this case the accuracy was 92% [6].

III. THE HITEG DATA GLOVE

For our experiment, to measure the hand's movements we adopted the so called HiTEg Glove V4, which consists of a supporting glove upon which 15 sensors are placed in correspondence of each distal interphalangeal joint, proximal interphalangeal joint and metacarpophalangeal joint. The glove was developed by the Health Involved Technical Engineering Group (HiTEg) at the University of Rome "Tor Vergata" [7], [8]. It is part of a system for measurement and reproduction in a virtual environment the static and dynamic postures of the hand. The HiTEg glove is integrated with a conditioning electronic circuitry, and a virtual representation on a computer screen by means of avatar. The conversion of the hand postures into electrical signals is due to bending and inertial sensors.

Figure 1: The Hiteg Data Glove

IV. ACQUISITION

Acquired electrical data are converted into voltage units in a digital 10-bit form, and a such resulting to values between 0 and 2^{10}-1 (1023), matching an interval of 0÷4.99V analogical values. The electric conversion is obtained by a voltage divider of known resistance placed in series with the unknown resistance represented by the sensor. However, what interests us is to acquire the angle of bending of each sensor which is precisely the data necessary to distinguish an act from another. Because the sensors are piezoresistive type their ohmic values vary depending on the angle of bending, so you must calculate the resistance of sensors operating the inverse formula of the following resistive voltage divider:
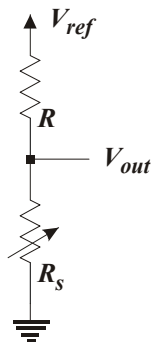
$$V_{out} = \frac{R_s}{R_s + R} V_{ref}$$

$$V_{out} = \frac{R_s}{R_s + R} V_{ref} \quad (1)$$

Figure 2: The voltage divider

The law characterizing the assumed angle of the sensor in relation to its ohmic value is not linear, as shown in figure 3. However, for simplicity reasons, in the practical phase of the project this trend is assumed linear, considering the fact that the margin of error is not relevant.
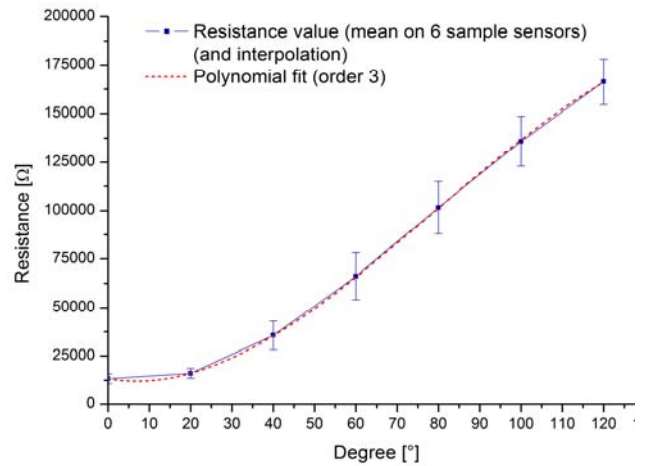

Figure 3: Sensor's resistance vs. bending angle

The bending angle of the sensor can be obtained from its resistance value through a calibration process, which store for each sensor its minimum ohmic value, which represents an angle of 0°, and its value resistive maximum, which is the maximum bending of the sensor.

The distal interphalangeal and metacarpophalangeal joints have a maximum bending angle of 90°, while the proximal interphalangeal may reach 120°. However, as a result of the limitation imposed by the architecture of the glove in the practical phase of the experiment, was considered, for all sensors, an angle in the range 0° - 90°.

V. THE SET OF GESTURES

The alphabet is composed of nineteen static gestures which are similar to most wide known languages of sign.
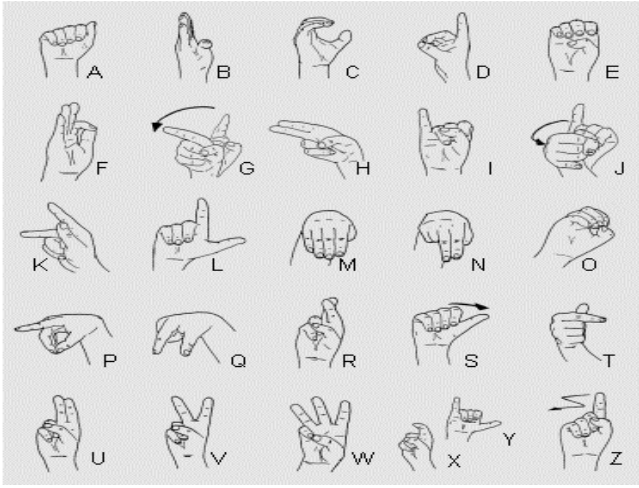
We cannot consider the whole LIS alphabet because some gestures are either dynamic or too ambiguous because of the hardware's limitations. However we have introduced new symbols that replaced those LIS gestures which are not represented in our set like "P","U","S","M","R", and finally "SPACE".

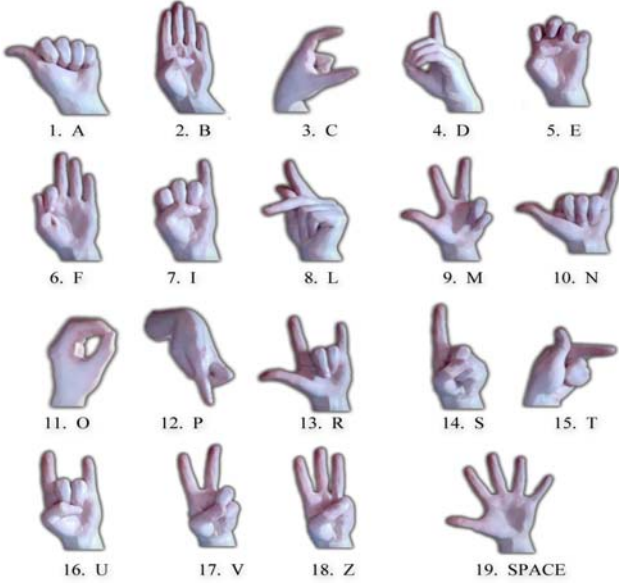The Figure 4a,b represent the real LIS gestures and the adopted one.

VI. CLASSIFICATION

To classify the acquired data we used three different algorithms: the Support Vector Machine (SVM), a Mahalanobis based classifier and the Euclidean based classifier.

In this section the three classifiers are briefly introduced. For detailed information and for a better understanding, see [9], [10], [11].

(a)

(b)

Figure 4: (a) The real LIS alphabet and (b) the used alphabet
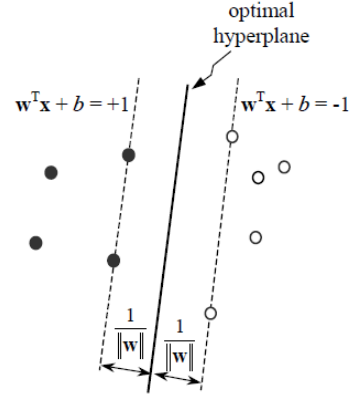

Figure 4: Optimal separating hyper-plane corresponding to the SVM solution. The support vectors lie on the dashed lines in the caption.

VII. SUPPORT VECTOR MACHINE

The aim of SVM is to find the hyperplane that maximize the separation between classes [9].

Let $(\overline{x}^k, y_k)$, $k=1,...,m$ represent the training examples for the classification problem; each example $\overline{x}^k \in \Re^N$ belongs to the class $y_k \in \{-1,+1\}$. Assuming linearly separable classes, there exists a separating hyper-plane such that

$$y_k (\overline{w}^T \overline{x}^k + b) > 0 \quad k=1,...,m \qquad (2)$$

The minimum distance between the data points and the separating hyper-plane is the margin of separation. The goal of a SVM is to maximize this margin. We can rescale the weights w and the bias b so that the constraints (2) can be rewritten as

$$y_k (\overline{w}^T \overline{x}^k + b) \geq 1 \quad k=1,...,m \qquad (3)$$

As a consequence, the margin of separation is $1/\|\overline{w}\|$ and maximization of the margin is equivalent to the minimization of the Euclidean norm of the weight vector $\overline{w}$. The corresponding weights and bias represent the optimal separating hyper-plane (Fig. 4). The data points $\overline{x}^k$ for which the constraints (3) are satisfied with the equality sign are called support vectors.

By means of Lagrange Multipliers we are able to consider only these vectors to find the optimal $\overline{w}$ and b. We use a Soft Margin SVM that introduces a tolerance to classification errors. The trade-off between the maximization of the margin and the minimization of the error is controlled by a constant *C*. In this context we have utilized a multiclass SVM which has two principal modality of learning: *One vs One* (1vs.1) and *One vs All* (1vs.A) [12].

We have adopted a Radial Basis Function kernel with the following parameter obtained through a validation test :

- $\gamma = 0,0000005$
- $C = 25$

VIII. DISTANCE BASED CLASSIFIERS

The following two classifiers are based on simple distance metrics.

The purpose of these classifiers is to infer the class of each sample taken from test set, by choosing the smallest distance from all the centroid in training set. The reader will note that in the following paragraphs, gestures are represented as a point in a multidimensional space.

We have decided to integrate these Classifiers with a k-nearest neighbor (k-NN) algorithm in order to classify only the closest training samples. In addition we redesigned these two classifiers using k-means procedure with the aim of avoid ambiguity among the classes of gestures.

A. EUCLIDEAN DISTANCE CLASSIFIER

The sample Euclidean distance classifier (EDC) has an easy mathematical model and it can be implemented without excessive difficulties.

Such a classifier needs a mathematical description of each hand posture. Consequently given a training set made up of n gestures, we calculate the correspondent centroid naming $c_k(f_1, f_2, ..., f_s)$, with $k=1, ... , n$ and s is the number of features taken into consideration. The centroid's features are defined as $f_i = \dfrac{1}{p} \sum\limits_{i=1}^{p} x_i$ with $i=1,$ .... , s while p is the number of samples for each class of gesture.

The next step is to finalize our training set by determining the radius of every gesture's distribution, and accordingly define the boundary of each class without ambiguity .

Since the classifier is a function that maps samples to the correspondent class, a gesture from test set can be evaluated by calculating distances between itself and the centroids of the whole training set.

The adopted metric is the Euclidean distance

$$d = \sqrt{(x_{f_1} - c_{f_1})^2 + (x_{f_2} - c_{f_2})^2 + ... + (x_{f_s} - c_{f_s})^2} \quad (4)$$

where d is the distance considered and x is gesture analyzed.

B. MAHALANOBIS DISTANCE CLASSIFIER

We also adopt Mahalanobis distance Classifier (MDC), because it fits well data which are widespread like Gaussian distribution in a s-dimension space. In addition, this metric emphasizes the correlations between variables, through which different patterns can be recognized.

Let x and c be respectively a generic sample from test set and the centroid of a specific class of gesture, both in the space R^s, then the distance d is defined in this way:

$$d(x,c) = \sqrt{(x-c)^T * CM^{-1} * (x-c)} \quad (5)$$

where CM is the covariance matrix of a specific class .

When the classifier makes use of Mahalanobis distance, it has to estimate covariance matrix of each class and at this point we can have an algebraic obstacle. In fact, computing the inverse of covariance matrix can be difficult if we are in the presence of singularity caused by the nature of the data distribution. Mathematically it is possible to avoid this complication by using Moore-Penrose pseudo-inverse [11] instead of normal inversion.

In any case, we can know on time if the previous computation leads to singularity by finding the condition number of CM. If condition number is near one, the matrix is well-conditioned and consequently we do not have problems. On the other hand, if it is zero, we fall into singularity.

IX. RESULTS AND DISCUSSION

We have collected data from users by applying these simple steps :
1. Wearing of glove
2. Evaluation of the glove adaptability
3. System Calibration
4. Recording of all the gestures of the Alphabet.

Thus, we call *session* the fulfilment of all previous steps.

It is important to take note of the shape and measurement of user's hand. As a consequence we synthesize those factors with a subjective parameter called glove adaptability which is empirically evaluated and can assume values between 1 and 3. For example if somebody has a smaller hand and finds a lot of difficulties on bending the sensors, we say that his glove adaptability is 1. Contrarily, in the optimal case, the glove adaptability must be 3.

Studying the motions of user's hand before executing the session is necessary since we have adopted for our experiment only one type of glove with a particular measurement.

We decided to take accuracy as a means of judgment.

This index is defined as the number of gestures correctly classified, over the number of the whole test set. The data set consists of 9 individuals(in the table, they will be numbered from S1 to S9) and based on experiments we are able to extract from them the samples for the training set and test set. The table that follows regards the accuracy of each classifier on a single person that performs one complete session:

TABLE 1: TEST ACCURACY (IN %) OF EACH CLASSIFIER ON EACH SUBJECT

	EDC	EDC-K	MDC	MDC-K	SVM 1vs.A	SVM 1vs.1
S1	84.0	84.0	73.6	63.0	94.7	94.7
S2	84.0	94.7	68.0	68.0	76.3	79.0
S3	84.5	94.7	89.5	89.5	92.0	89.5
S4	100.0	100.0	94.7	89.5	100.0	100.0
S5	89.5	94.7	89.5	84.0	94.7	94.7
S6	94.7	94.7	68.0	79.0	81.5	73.7
S7	84.0	84.0	73.6	79.0	84.2	84.0
S8	100.0	94.7	73.6	79.0	94.7	94.7
S9	84.0	84.0	89.0	84.0	84.0	94.7

Nine mentally healthy subjects (six males and three females) 25-40 aged were involved in the experiments.

The following tables (Tables 2 and 3) report the mean accuracies of each classifiers in the case where the training subjects and the test subjects are the same (Table 2), and where training subjects and test subjects are different (Table 3).

The latter case is very important to take into consideration: in fact, if the obtained percentage is high, it means that the software can be trained by a person before and then used by other people who did not participate in the learning phase, making them to save a lot of training time.

In the first case the training and test sets were divided as follows:

Training set: two sessions from S1, S2, S3, S4, S5
Test set: one session from S1, S3, S4, S5

TABLE 2: ACCURACY OF EACH CLASSIFIER ON SUBJECTS THAT BELONG BOTH TO TRAINING SET AND TEST SET

EDC	EDC-K	MDC	MDC-K	SVM 1vs1	SVM 1vsA
89.0%	95.0%	66.0%	63.0%	92.0%	92.0%

In the second case the training and test sets were divided as follows:

Training set: two session from S1, S2, S3, S4, S5
Test set: one session from S6, S7, S8, S9

TABLE 3: ACCURACY OF EACH CLASSIFIER ON SUBJECTS THAT BELONG TO THE TEST SET BUT NOT TO TRAINING SET

EDC	EDC-K	MDC	MDC-K	SVM 1vs1	SVM 1vsA
91.00%	79.00%	39.00%	43.00%	87.00%	89.50%

From the table we can notice that EDC and SVM classifiers fit well to our proposals in terms of accuracy.
The results emphasized that there is no difference between the two versions of SVM (SVM one versus one and SVM one versus All).
It can also be noticed that Euclidean classifier works better than Mahalanobis in both cases and the reason is related to the distribution's nature of our data set. In fact we can deduce from the result that samples of each class of gestures are located in the s-dimension space near the corresponding centroid, forming an hyper-spherical distribution. This could be expected since we have considered only static gestures.
Regarding the K-Means, it turned out that while it does not so much affect the MDC and the SVM, it increases the accuracy of EDC when used on subjects that belong both to training and test set; on the other hand it drastically decreases the performance when the subjects from the training set and test are different.

X. CONCLUSIONS

In this paper we compared three different classifiers for static hand gestures.
To do these experiments, we started from different assumptions. Since we studied static gestures, we did not take into consideration values from accelerometers. The law which characterizes bending angles is assumed linear (see *The hiteg data glove*). Finally, 90 degrees is considered as the highest value of bending for each finger.
It turned out that the best classifiers were SVM and EBC, with an outstanding good accuracy with peaks above 90%.

It is also important to notice that this kind of classification problem is independent from who recorded the training set, hence the classifiers can be trained in a first stage and then used by other users without re-training the system.
This type of project is very important in social field. The original idea was to create a system that helps the deaf to communicate with other people. But this study has many more applications. For example, it is possible to use our system to substitute keyboard as a way of writing or to associate gestures to commands in order to control external devices (such as mechanical arms, robots, etc.) or even use the data glove as a controller for gaming applications.
Our next step will be the recognition of dynamic gestures with the proposal of identifying not only symbols from alphabet but also concepts expressed in sign language.

REFERENCES

[1] Y. Wu and T. S. Huang, "Vision Based Gesture Recognition A Review", *International Gesture Workshop*, GW 99 France, March 1999.

[2] S. Mitra and T. Acharya, "Gesture Recognition: A Survey", *IEEE Transactions on Systems, Man and Cybernetics*, part C vol. 37, no. 3, May 2007, pp. 311-324.

[3] T. E. Starner and A. Pentland, "Visual Recognition of American Sign Language using Hidden Markov Models", *Intl'l Workshop Automatic Face and Gesture Recognition*, IEEE Press, 1995, pp. 189-194.

[4] T. Takahashi and F. Kishino, "Hand Gesture Coding based on Experiments using a Hand Gesture Interface Device", *SIGCHI Bulletin*, vol. 23, no. 2, 1991, pp. 67-74.

[5] K. Jong-Sung, W. Jang and A. Bien "A Dynamic Gesture Recognition System for the Korean Sign Language (KSL)", *IEEE Transactions on Systems, Man and Cybernetics*, part B vol. 26, no. 2, April 1996, pp. 354-359.

[6] D. Xu, "A Neural Network Approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG", *18th International Conference on Pattern Recognition*, March 2006, pp. 519-522.

[7] G. Costantini, G. Saggio and M. Todisco, "A Glove Based Adaptive Sensor Interface for Live Musical Performances", *1st International Conference on Sensor Device Technologies and Applications*, July 2010.

[8] G. Orengo, L. Giovannini, G. Latessa, G. Saggio and F. Giannini, Characterization of Piezoresistive Sensors for Goniometric Glove in Hand Prostheses", *Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology*, May 2009, pp. 684-687.

[9] V. N. Vapnik, "The Nature of Statistical Learning Theory", *Springer-Verlag* 1999.

[10] P. C. Mahalanobis, "On the Generalised Distance in Statistics", *Proceedings of the National Institute of Sciences of India*, 1936.

[11] R. Penrose, "A Generalized Inverse for Matrices", *Proc. Cambridge Phil. Soc. 51*, 1955, pp. 406-413.

[12] Chih-Wei Hsu, Chih-Jen Lin, "A comparison of Methods for Multi-class Support Vector Machine", *IEEE Transaction on Neural Networks*, vol. 13, no. 2, March 2002, pp. 415-425.