Fuzzy Discriminant Analysis based Feature Projection in Myoelectric Control

Rami N. Khushaba, Student *Member, IEEE*, Adel Al-Jumaily, *Member, IEEE*, and Ahmed Al-Ani

Abstract— The myoelectric signal (MES) from human muscles is usually utilized as an input to the controller of a multifunction prosthetic hand. In such a system, a pattern recognition approach is usually employed to discriminate between the MES from different classes. Since the MES is recorded using multi channels, the feature vector size can become very large. In order to reduce the computational cost and enhance the generalization capability of the classifier, a dimensionality reduction method is needed to identify an informative moderate size feature set. This paper proposes a novel feature projection technique based on a combination of Fisher's Linear Discriminant Analysis (LDA), and Fuzzy Logic. The new method, called FLDA, assigns different membership degrees to the data points thus reducing the effect of overlapping points in the discrimination process. Furthermore, the concept of Mutual Information (MI) is introduced in the fuzzy memberships in order to assign weights to the features (attributes) according to their contribution to the discrimination process. The FLDA method is tested on a seven classes MES dataset and compared with other feature projection techniques proving its superiority.

I. INTRODUCTION

THE myoelectric signal (MES) also known as the Electromyogram (EMG), is a one dimensional non-stationary signal that carries the distinct signature of the voluntary intent of the central nervous system. It is usually measured in a noninvasive scheme utilizing a set of surface electrodes mounted on the human forearm. The most important application of the MES is its use in controlling prosthetic devices functioning as artificial alternatives to missing limbs. Advances in myoelectric signals studies revealed that the MES exhibits different temporal structure for different kinds of the arm movements. This in turn facilitated the use of a pattern recognition based myoelectric control strategies for prosthesis control. To this end, a wide set of pattern recognition methods were proposed in the literature to produce a computationally efficient and accurate MES recognition systems [1].

Although many of the proposed systems achieved good classification accuracies, but the actual success of such systems is measured by means of clinical availability. The most important factor that limits the clinical availability is the control strategy itself. One of the significant advantages

All the authors are within the Mechatronics and Intelligent Systems Group, University of Technology, Sydney. P. O. Box: 123, Broadway 2007, Sydney, Australia.

of pattern recognition based myoelectric control strategies is that it represents an intention based control, i.e., a hand-free control scheme [1]. A common issue with most amputees is that the nerves underlying the skin can still be innervated and are thus subject to voluntary control based on the user's intention. On the other hand, there are currently two very important limitations in the available myoelectric control strategies [2]. The first limitation is the huge amount of myoelectric data required to be processed and used as a control input for the prosthesis. This is mainly caused by the fact that the MES detected at different locations over the same muscle may have significantly different amplitudes [3]. This in turn necessitates the use of a larger set of electrodes to capture the complete muscle activity, which usually introduces a *time delay* in the myoelectric control system that hinders the development of a continuous control. The second limitation is imposed by the first one, as in order to reduce the computational cost of such a system, several techniques are utilized to reduce the size of the resultant features vector. But the performance of many of the available techniques is far from optimal, thus *decreasing the classification accuracy* of such a system.

The goal of this paper is to produce a new feature projection method based on a combination of Fisher's linear discriminant analysis (LDA) and fuzzy logic. The proposed feature projection method, termed as FLDA, aims to reduce the dimensionality of the extracted feature set and cluster features, such that the classification accuracy is improved. To achieve these goals, the proposed method incorporates fuzzy memberships to reduce the impact of outliers on the formed cluster's structure. A possibilistic fuzzy clustering approach is utilized to calculate the membership values. Furthermore, the concept of mutual information is introduced in the calculations of the memberships thus assigning certain importance values to the contribution of each feature.

This paper is structured as follows: Section II presents the background and related work. Section III explains the proposed methodology. The experimental results are given in section IV. Finally the conclusion is given in section V.

II. BACKGROUND AND RELATED WORK

Features are computed from the preprocessed MES using a sliding analysis window, where a single feature vector is produced from each analysis window. A wide variety of feature extraction techniques were utilized in the literature

with the MES signal ranging from simple time domain analysis to the time-frequency analysis methods [1]. Advances in myoelectric control problems revealed that the simple combination of Hudgins's time domain (TD) features and the autoregressive (AR) features had achieved good classification accuracies while at the same time being computationally efficient [4, 5]. However, an argument still exist in the literature that systems developed based on these simple TD features may not present high success rates [6]. This is due to the fact that the time domain methods treat the MES as a stationary signal, while in reality it is not.

In order to capture the complete muscle activity, a multi channel approach is usually utilized when measuring the MES signal. However, this will increase the number of extracted features and hence it will increase the learning parameters of the classifier and may degrade its convergence [7]. In order to reduce the system's computational complexity and enhance the generalization capability of the classifier a dimensionality reduction technique will be required.

A straight forward solution to these problems is to project the data onto low-dimensional subspaces to extract the most significant features. Many feature projections techniques were used in myoelectric control with the aim to produce a statistically uncorrelated or independent feature set, a desirable goal in any pattern recognition system. The first study on the effect of different dimensionality reduction techniques in the MES pattern recognition problem was conducted by Englehart [8]. Englehart's study revealed that the combination of wavelet packet transform based features with principal component analysis (PCA), a well-known feature projection method, gave powerful performance when compared with other methods [9]. Some of the proposed feature projection methods in myoelectric control include the combination of PCA and a self organizing feature map (SOFM), proving better results than PCA alone [7], linear discriminant analysis (LDA) based feature projection, proving better results than PCA with SOFM [10], and the uncorrelated linear discriminant analysis (ULDA) that requires the reduced features to be statistically uncorrelated with one another [5], a more powerful version of LDA.

In order to produce a fair comparison, two feature sets will be extracted in this paper. The first set consists of the combination of TD and AR features, termed TDAR. The second feature set includes the mean value of the squares of the wavelet coefficients extracted by using a Symmlet family of wavelets. The proposed FLDA feature projection method will be compared with ULDA, and PCA using both feature extraction techniques mentioned.

The theory behind the proposed method is introduced next.

III. METHODOLOGY

Consider a classification problem with c-classes, in which the data set of labeled training samples is given as:

$$S = \{(x_1, y_1), \ldots, (x_l, y_l)\} \subseteq (X \times Y)^l \qquad \ldots (1)$$

Where X is the input space and Y is the output space. $X \subseteq R^n, Y \subseteq R^n$, and l is the number of samples. Each training point x_i originally belongs to one of the c classes and is given a label $y_i \in \{1, 2, 3, \ldots, c\}$ for $i=1, \ldots, l$. The goal is to find an optimal hyper-plane using the training samples that can recognize the test points, i.e., the classifier will have a good generalization capability. In FLDA each point, x_i, belongs to each of the c classes with a certain membership. The fuzzy within class scatter matrix S_W, fuzzy between class scatter matrix S_B, and the fuzzy total class scatter matrix S_T are given as follows:

$$S_W = \sum_{i=1}^{c} \sum_{k=1}^{l} u_{ik}^m (x_k - v_i)(x_k - v_i)^T \qquad \ldots (2)$$

$$S_B = \sum_{i=1}^{c} \sum_{k=1}^{l} u_{ik}^m (v_i - \bar{x})(v_i - \bar{x})^T \qquad \ldots (3)$$

$$S_T = \sum_{i=1}^{c} \sum_{k=1}^{l} u_{ik}^m (x_k - \bar{x})(x_k - \bar{x})^T \qquad \ldots (4)$$

Where u_{ik} is the membership of pattern k in class i, v_i is the mean of the patterns belonging to class i, and $\bar{x}$ is the mean of the training samples given by:

$$\bar{x} = \frac{1}{l} \sum_{k=1}^{l} x_k \qquad \ldots (5)$$

The values of the membership function u_{ij}^m are calculated using a possibilistic fuzzy clustering approach. The cost function of the possibilistic clustering approach is given by the following equation [11]:

$$J(\theta, U) = \sum_{k=1}^{l} \sum_{i=1}^{c} u_{ik}^m d(x_k - \theta_i) + \sum_{i=1}^{c} \eta_i \sum_{k=1}^{l} (1 - u_{ik})^m \qquad \ldots (6)$$

where η_i are suitably chosen positive constants, and θ_i are the values of the clusters centers.

The first term in Eq. (6) is the same objective function used in probabilistic clustering approach, while the second term is added to reduce the effect of outliers. In order to find the membership values then we first need the values of the clusters centers. A direct way would be to differentiate Eq. (6) with respect to θ, but this in turn would remove the second term leaving only the first term. A general look at the first term of Eq. (6) reveals that this is the classical within class scatter matrix S_W. Thus applying the values of the clusters means ensures that the objective function would settle at a global optimum. In order to compute the membership values one can simply differentiate the resultant function with respect to u as in Eq. (7) and (8).

The final issue to mention here is that a weighted difference is applied in Eq. (8). Also the values of η_i were chosen according to the same approach mentioned in [12].

$$\frac{\partial J(\theta, U)}{\partial u_{ik}} = m\, u_{ik}^{m-1}\, d(x_k, v_i) - m\, \eta_i\, (1 - u_{ik})^{m-1} = 0 \quad \ldots(7)$$

Thus
$$u_{ik} = \frac{1}{1 + \left(\dfrac{d(x_k, v_i)}{\eta_i}\right)^{\frac{1}{m-1}}} \quad \ldots(8)$$

The idea behind using a weighted difference in the membership calculation is that certain features are more significant that others. In simple words, treating all features equally can have a negative impact on the overall performance. A relevant feature will produce a better subset, and hence improve the membership estimation, while an irrelevant feature may degrade the performance. The weighting scheme proposed in this paper is based on the concept of mutual information (MI). The MI measures the arbitrary dependencies between random variables and is suitable for assessing the information content. Fano [13] has shown that maximizing the MI between transformed data and the desired target achieves the lowest probability of error. The MI between a given feature subset L and class labels C, is measured by the following formula:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log\left(\frac{p(x, y)}{p_1(x)\, p_2(y)}\right), \quad \ldots(9)$$

where $p(x,y)$ is the joint probability distribution function of X and Y, and $p_1(x)$ and $p_2(x)$ are the marginal probability distribution functions of X and Y.

After computing all the variables, FLDA work to find the vector G that would maximize the ratio of the between class scatter matrix to the within class scatter matrix by solving the following equation

$$G = \arg\max_{G} trace\left((G^T S_W G)^{-1} G^T S_B G\right) \quad \ldots(10)$$

The solution can be readily computed by applying an eign-decomposition on $S_W^{-1} S_B$, provided that the within class scatter matrix S_W is nonsingular. Since the rank of the between class scatter matrix is bounded from above by $c - 1$, there are at most $c - 1$ discriminant vectors by FLDA. An intrinsic limitation of both LDA and FLDA is that the objective function used with both techniques requires the total scatter matrix to be nonsingular, were as in some real problems the scatter matrix can be singular. To address this problem, singular values decomposition was employed, thus resembling the implementation of ULDA but with fuzzy memberships.

IV. EXPERIMENTAL RESULTS

The MES dataset used in this research is the same one that was originally used by Chan *et al* [5]. Eight channels of surface MES were recorded from the right arm of thirty normally limbed subjects (twelve males and eighteen females). Each subject underwent four sessions, with one to two days separation between sessions. Each session consisted of six trials. Seven distinct limb motions were used, hand open (HO), hand close (HC), supination (S), pronation (P), wrist flexion (WF), wrist extension (WE), and rest state (R). The positions of the electrodes are as shown in Fig.1. As in the original research of Chan *et al* [5], data from only session four were used. Data from the first two trials were used as training set and data from the remaining four trials were used for testing.

Two sets of features were extracted from the dataset in order to test the performance of the proposed method with different feature extraction techniques. The first set of extracted features included a combination of the first four autoregressive coefficients and the root mean square value as the feature vector (total of 40 features = 8 channels × 5 features/channel). This feature set was referred to as the TDAR feature set. The analysis window size was 256 ms which were spaced 128 ms apart for training data and 32 ms apart for testing data. Data that were 256 ms before or after a change in limb motion were removed from the training set to avoid transitional data. The second feature set extracted included the mean of the square values of the wavelet coefficients using a Symmlet wavelet family with five levels of decomposition (total of 48 feature = 8 channels × 6 features/channel).

The classification results for both feature extraction techniques are shown are shown in Fig.2, employing LDA as a classifier. The results represent the averaged classification accuracies across thirty subjects. Both feature sets were reduced in dimensionality with FLDA, ULDA, and PCA. It is obvious from these results that the proposed FLDA outperformed all other methods of feature projection. The performance of ULDA was not far from that of FLDA, and both are clearly better than PCA. A quick look reveals that the WT features achieved higher accuracies than that achieved by the TDAR features. This is due to the fact that the wavelet transform contains information about the MES signal from both the time and the frequency domains.
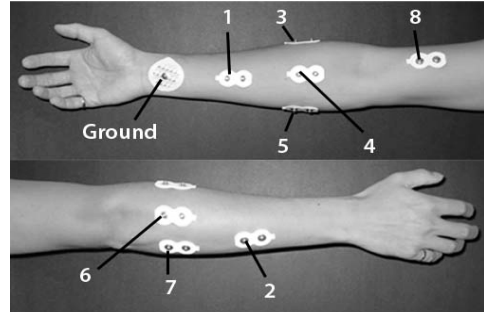

Fig.1 Electrodes placement on the right arm

The reason of the lower performance achieved with the TDAR feature is actually due to the quality of the features themselves. Although the performance of the TDAR feature set is lower than that of the WT feature set, but it should be pointed out that the computational cost for the TDAR feature set is less than that of the WT feature set. Thus the choice of one feature set over another can be left to the designer.

In order to provide a rigorous validation or comparison with existing techniques for dimensionality reduction, the confusion matrix for all the subjects was computed for the WT feature set. A plot of the diagonal values of the confusion matrices (class-wise classification accuracies) is shown in Fig.3 for the three feature projection techniques.

The results in Fig.3 indicate the significance of including the membership values of each pattern in the calculations of the scatter matrices. In this case the patterns that belong to different classes with same memberships will have a smaller effect in determining the structure of the clusters formed. While at the same time the pattern that are concentrated in a certain cluster center will have a higher effect in forming good clusters.

V. CONCLUSION

A new fuzzy discriminant analysis feature projection technique was introduced in the field of myoelectric control. The proposed FLDA was built based on a combination of Fisher's linear discriminant analysis and fuzzy memberships. The advantages of the proposed method were obvious in achieving higher classification accuracies. It was also shown that assigning different membership value to the patterns did actually help in forming good clusters. This in turn reduced the effect of outlier points on the discrimination process. A limitation of the proposed method was also pointed out with the technique used to overcome such a limitation. As advancement to the proposed method, we are currently utilizing an optimization technique with FLDA to further enhance its performance.

Fig.2 Average classification accuracies with different feature sets and dimensionality reduction techniques.

Fig.3 Average confusion matrix across thirty subjects

REFERENCES

[1] M. Asghari Oskoei and H. Hu, "Myoelectric control systems--A survey," *Biomedical Signal Processing and Control*, vol. 2, pp. 275-294, 2007.

[2] R. N. Khushaba, A. Al-Jumaily, and A. Al-Ani, "Novel feature extraction method based on fuzzy entropy and wavelet packet transform for myoelectric Control," presented at ISCIT '07. International Symposium on Communications and Information Technologies., 2007.

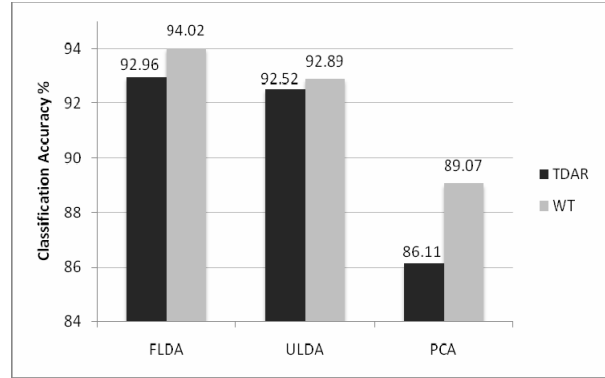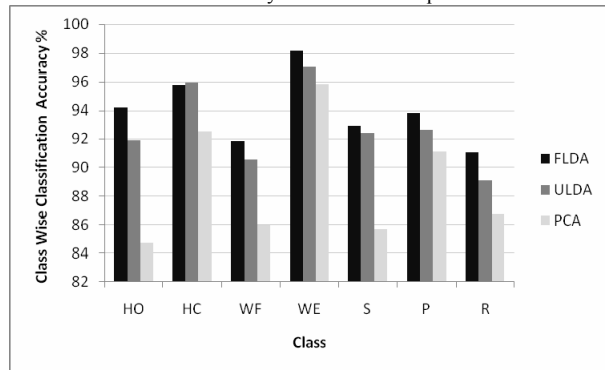[3] R. Merletti, A. Bottin, C. Cescon, D. Farina, M. Gazzoni, S. Martina, L. Mesin, M. Pozzo, A. Rainoldi, and P. Enck, "Multichannel surface EMG for the non-invasive assessment of the anal sphincter muscle," *Digestion*, vol. 69, pp. 112-122, 2004.

[4] L. Hargrove, K. Englehart, and B. Hudgins, "A comparison of surface and intramuscular myoelectric signal classification," *IEEE Transactions on Biomedical Engineering*, vol. 54, pp. 847-853, 2007.

[5] A. D. C. Chan and G. C. Green, "Myoelectric control development toolbox," presented at accepted to the 30th Conference of the Canadian Medical & Biological Engineering Society, Toronto, ON, 2007.

[6] C. Jun-Uk, M. Inhyuk, K. Shin-Ki, M.-S. Mun, and A. M.-S. Mun, "Control of multifunction myoelectric hand using a real-time EMG pattern recognition," presented at IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). , 2005.

[7] J. U. Chu, I. Moon, and M. S. Mun, "A real-time EMG pattern recognition system based on linear-nonlinear feature projection for a multifunction myoelectric hand," *IEEE Transactions on Biomedical Engineering*, vol. 53, pp. 2232-2239, 2006.

[8] K. Englehart, "Signal representation for classification of the transient myoelectric signal " in *Electrical and Computer Engineering Department.*, vol. PhD Dissertation: University of New Brunswick, 1998.

[9] K. Englehart, B. Hudgins, P. Parker, and M. Stevenson, "Time-frequency representation for classification of the transient myoelectric signal," in *the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 5, 1998, pp. 2627-2630.

[10] J. U. Chu, I. Moon, and M. S. Mun, "A supervised feature projection for real-time multifunction myoelectric hand control," presented at The 28th IEEE EMBS Annual International Conference, New York City, USA, 2006.

[11] J. V. D. Oliveira and W. Pedrycz, "A comprehensive, coherent, and in depth presentation of the state of the art in fuzzy clustering," John Wiley & Sons Ltd, 2007.

[12] X.-H. Wu and J.-J. Zhou, "Fuzzy discriminant analysis with kernel methods," *Pattern Recognition*, vol. 39, pp. 2236-2239, 2006.

[13] R. M. Fano, Transmission of information; a statistical theory of communications: Cambridge, Mass., M.I.T. Press, 1961.