IDIAP RESEARCH REPORT

IMPROVING CONTROL OF DEXTEROUS HAND PROSTHESES USING ADAPTIVE LEARNING

Tatiana Tommasi Francesco Orabona
Claudio Castellini Barbara Caputo

Idiap-RR-07-2012

JANUARY 2012

Centre du Parc, Rue Marconi 19, P.O. Box 592, CH - 1920 Martigny
T +41 27 721 77 11 F +41 27 721 77 12 info@idiap.ch www.idiap.ch

Improving Control of Dexterous Hand Prostheses Using Adaptive Learning

Tatiana Tommasi, Francesco Orabona, Claudio Castellini and Barbara Caputo

***Abstract*—At the time of writing, highly dexterous hand prostheses are being manufactured and, to some extent, marketed. This means wearable, implantable mechanical hands with many independently controllable degrees of freedom, e.g. finger flexion and thumb rotation. Still, control by the patient is an open issue, and the most promising way ahead is probably machine learning applied to surface electromyography (sEMG). Researchers have mainly concentrated so far on improving the accuracy of sEMG classification and/or regression; but in general, a finer control implies longer and harder training times. A more natural form of control might shorten the time a patient requires to learn how to use the prosthesis, but the machine training time will inevitably be longer.**

In this work we propose a general method to re-use past experience, in the form of models synthesised from previous users, to boost the adaptivity of the prosthesis and dramatically shorten the training time. Extensive tests on a database recorded from 10 healthy subjects in controlled and non-controlled conditions reveal that the method is highly effective.

***Index Terms*—learning and adaptive systems, prosthetics, electromyography, brain-computer interfaces**

I. Introduction

In the prosthetics / rehabilitation robotics community it is generally understood nowadays (see, e.g., [1], [2], [3]) that advanced hand prostheses need adequate control schema, where by "adequate" one means accurate, reliable and easy to use from the point of view of the subject who is wearing them. In particular, polyarticulate, force-controlled hand prostheses are currently appearing in the clinical setting; e.g., the *BeBionic* hand by RSL Steeper (www.bebionic.com), Vincent Systems' *Vincent hand* (www.handprothese.de) and the *i-LIMB* by Touch Bionics (www.touchbionics.com, see Figure 1). These hands enjoy a much higher dexterity than before (although not yet comparable with non-prosthetic mechanical hands), but still, control by the patient is poor. In particular, hand / finger postures are achieved via complex sequences of muscle contraction impulses that one must get acquainted with and remember. Together with excessive weight and low reliability, lack of control is deemed to be one of the main reasons why 30% to 50% of upper-limb amputees do not use their prosthesis regularly [4].

To overcome at least one aspect of this rejection, a more natural form of control has been individuated and studied since two decades; namely, an old method, surface electromyography (sEMG), revamped by the application of machine learning techniques to its signal. Actually, sEMG is in the clinical usage since the 60s, typically employing two electrodes and a threshold detector [5], [6], [7]. On the other hand, thanks to adequate arrays of electrodes and statistical classification and / or regression techniques (e.g., support vector machines, linear discriminants, neural networks) we are now able, at least in principle, to detect what the patient wants to do and to enforce it. This is the meaning of "natural control". Machine learning is highly adaptive, eliminating the need for the stump analysis phase and allowing natural feed-forward control. Recent results on amputees indicate that even long-term patients can generate precise residual activity, actually as precise as that generated by intact subjects (e.g. [8]). The potentiality of controlling a dexterous hand prosthesis to a degree of finesse unknown so far is therefore at hand. The use of supervised learning ensures that a well trained prosthesis will act, in principle, as the patient's will dictates.

Fig. 1. Dexterous hand prostheses: (left to right) RSL Steeper's *BeBionic* (reproduced from www.bebionic.com), Vincent Systems' *Vincent hand* (www.handprothese.de) and Touch Bionics's *i-LIMB Ultra* (www.touchbionics.com).

In this paper we concentrate upon a slightly different aspect of machine learning applied to hand prostheses control, namely, we try to reduce the *training time*, i.e. the time required to perform the adaptation of the prosthesis itself to the patient. The use of supervised machine learning implies that a (usually large) set of labeled samples must be acquired from each patient anew, in order to *train* the system; this process is usually long and not guaranteed to necessarily produce a good sEMG-to-hand-configuration model. Anatomical similarity among humans intuitively suggests that good statistical models built in the past might be proficiently reused when training a prosthesis for a new patient; but this idea cannot be naïvely enforced, as shown at least in [9], where cross-subject analysis (i.e., using a model trained on a subject to do prediction on a new subject) is performed with poor results.

T. Tommasi and B. Caputo are with the Idiap Research Institute, Martigny, Switzerland, and the École Polytechnique Fédérale de Lausanne, Switzerland. Email: tatiana.tommasi@idiap.ch, barbara.caputo@idiap.ch

Francesco Orabona, Toyota Technological Institute at Chicago, Illinois, USA. Email: francesco@orabona.com

Claudio Castellini is with the Institute of Robotics and Mechatronics, German Aerospace Research Center, Oberpfaffenhofen, Germany. Email: claudio.castellini@dlr.de

To enforce a more refined approach to the idea, we present hereby the results of extensive experiments in which *adaptive learning* has been applied to the problem in order to boost the training phase of a hand prosthesis by reusing previous experience. We build on our own previous work [10] which proposed a principled method to choose the pre-trained models for adaptation, and the degree of *closeness* between each of them and the new target task. This approach was based on an estimate of the model generalization ability through the leave-one-out error. Here we improve the original method into two key aspects: (1) we constrain the new model, learned online by the patient, to be close to a subset of pre-trained models, as opposed to a single one as in [10], stored in the memory of the prosthesis; (2) the learning process to define from whom and how much to adapt is now defined through a convex optimization problem while in [10] this goal was achieved solving a non-convex optimization task prone to local minima issues. This translates into a consistent bootstrapping of the control abilities of the subject, which can therefore acquire control of the device within a time range of at least one order of magnitude lower than what would be achieved by using the data of the patient only.

We test our method on the database already described in [9], [10], consisting of sEMG, posture and force signals gathered from 10 intact subjects in various (controlled, non-controlled) laboratory situations. The benefits are apparent: when models bootstrapped in such a way are tested on subjects so far unseen, and in non-controlled conditions, the error rates are significantly lower than in the non-bootstrapped case, since the beginning and *along the whole process*. The perspective is that of shipping a pre-trained prosthesis which would very quickly adapt to the patient, with the effect of enabling him / her to a higher comfort and aid during daily-life activities.

The paper is organised as follows: after reviewing related work, in Sections II and III we present the method. Section IV describes the database used, while Section V shows and discusses the results. Lastly, Section VI contains conclusions and ideas for the future work.

A. Related work

1) Using sEMG for hand prostheses: Research in the field of machine learning applied to sEMG is, to date, deep and quite mature, at least as far as classification of hand postures and prediction of required force is concerned. Surface EMG detects muscle unit activation potentials, in the typical case quasi-linearly related to the force exerted by the muscle to which it is applied. In the more specific case of hand prostheses, several electrodes are applied to the forearm (or stump) while the subject reaches specific hand configurations (postures) and/or grabs a force sensor. The raw sEMG signal is then preprocessed (filtered, rectified, subsampled) and fed, together with force values and labels denoting the postures, to a supervised machine learning method. Hand postures are classified accordingly, and the force applied is predicted using a regression scheme. The two things can happen simultaneously [11]. Up to 12 hand postures [12] have been classified with acceptable accuracy, and there are strong hints [8], [13], [12], [14] that most trans-radial amputees may achieve similar performance. Almost comprehensive surveys can be found in [2], [3] and the most recent results at the time of writing are probably those exposed in [15], [16], [17] and [18]. It seems then, that any general-purpose method to speed up the training time / aid the collection of training data is highly desirable. To the best of our knowledge, [10] and [19] presented the only two attempts of using model adaptation to boost learning on sEMG data of different subjects. In [10] previously trained models are exploited as starting point when learning on a new subject. In [19] reweighted samples from multiple source subjects are combined to the target subject samples.

2) Adaptive learning: In the last years, different machine learning techniques for domain adaptation [20] have gathered attention in natural language processing [21], [22], computer vision [23] and sentiment classification [24]. The goal of domain adaptation is to overcome the systematic bias that occur between different sets for the same learning task. More precisely, if we call (X, Y) and $(\tilde{X}, \tilde{Y})$ the data and corresponding labels coming from a source and a target set, domain adaptation addresses the case where the conditional distribution of labels are similar $P(Y|X) \simeq P(\tilde{Y}|\tilde{Y})$ but the corresponding marginal distributions $P(X)$ and $P(\tilde{X})$ differ. By applying domain adaptation, data collected in different domains can be used together (source + target) or it is possible to leverage on pre-trained models built on rich training sets (source) when facing the same problem in a new domain with few available samples (target).

Many adaptive methods have been compared and benchmarked in [22], however most of them are computationally inefficient because it is necessary to retrain each time over old and new data. A domain adaptation method that follows this line has been presented in [19]. The source samples are reweighted on the basis of both the marginal and the conditional probability differences between each source and the target task. When learning the final classifier on the whole set of samples an additional weighting factor is added to evaluate the real relevance of each source with respect to the target task. The sensibility of the method to this parameter is evaluated empirically, but how to choose it is left as an open problem.

An approach that does not use re-training, based on SVM has been proposed in [25]. However this technique does not take properly into account the possibility that the known model can be too different with respect to the new one due to high variability among the domains. In [10] we built on [25] and overcame the described drawback introducing a principled method which choose the best prior knowledge to use and how much to rely on it. We propose here to enlarge this approach and make it able to combine multiple prior knowledge models at the same time with the aim to exploit at the best all the available information.

II. Defining the Adaptive Model

In this section we describe the mathematical framework at the basis of our adaptive learning method. We first introduce the basic notation (Section II-A), then we present our algorithm for online model adaptation from the best known subject

(Section II-B) and how to enlarge it to exploit multiple known subjects (Sections II-C).

A. Background

Assume $\mathbf{x}_i \in \mathbb{R}^m$ is an input vector and $y_i \in \mathbb{R}$ is its associated output. Given a set $\{\mathbf{x}_i, y_i\}_{i=1}^l$ of samples drawn from an unknown probability distribution, we want to find a function $f(\mathbf{x})$ such that it determines the best corresponding y for any future sample $\mathbf{x}$. This is a general framework that includes both regression and classification. The problem can be solved in various ways. Here we will use kernel methods and in particular Least-Squares Support Vector Machines (LS-SVM, [26]). In LS-SVM the function $f(\mathbf{x})$ is built as a linear model $\mathbf{w}\cdot\phi(\mathbf{x})+b$, where $\phi(\cdot)$ is a non-linear function mapping input samples to a high-dimensional (possibly infinite-dimensional) Hilbert space called *feature space*. Rather than being directly specified, the feature space is usually induced by a *kernel function* $K(\mathbf{x}, \mathbf{x}')$ which evaluates the inner product of two samples in the feature space itself, i.e. $K(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})\cdot\phi(\mathbf{x}')$. A common kernel function is the Gaussian kernel:

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma||x - x'||^2) \tag{1}$$

that will be used in all our experiments.

The parameters of the linear model, $\mathbf{w}$ and b, are found by minimizing a regularized least-squares loss function [26]. This approach is similar to the well-known formulation of Support Vector Machines (SVMs), the difference being that the loss function is the square loss. While this does not induce a sparse solution, it makes it possible to write the leave-one-out error in closed form and with a negligible additional computational cost [27]. This is known to be approximately an unbiased estimator of the classifier generalization error [28]. This property is useful to find the best parameters for learning (e.g. γ in (1)) and it will be used in our adaptation method. Note that we use the same general formulation to solve both regression and classification problems.

B. Model Adaptation from the Best Subject

Let us assume we have N pre-trained models stored in memory, trained off-line on data acquired on N different subjects. When the prosthetic hand starts to be used by subject $N+1$, the system begins to acquire new data. Given the differences among the subjects' arms and as well in the placement of the electrodes, these new data will belong to a new probability distribution, in general different from the N previously modeled and stored. Still, as all subjects perform the same grasp types, it is reasonable to expect that the new distribution will be *close* to at least one of those already modeled; then, it should be possible to use one of the pre-trained model as a *starting point* for training using the new data. We expect that, by doing so, learning should be faster than using the new data alone. To solve this problem we generalize the framework for adaptation proposed in [25] for SVMs: the basic idea is to slightly change the regularization term of the SVM cost functional, so that the solution will be close to the pre-trained one. The optimization problem is [25]:

$$\begin{aligned} &\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w} - \mathbf{w}'\|^2 + C\sum_{i=1}^{l}\xi_i \\ &\text{subject to } \ \xi_i \geq 0, \ \ y_i(\mathbf{w}\cdot\phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \end{aligned} \tag{2}$$

where $\mathbf{w}'$ is a pre-trained model. In order to tune the closeness of $\mathbf{w}$ to $\mathbf{w}'$, we introduce a scaling factor β weighing the pre-trained model; also, we use the square loss and therefore resort to the LS-SVM formulation. In this way the leave-one-out error can be evaluated in closed form, enabling automatic tuning of β. The optimization problem reads now like this [10]:

$$\begin{aligned} &\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w} - \beta\mathbf{w}'\|^2 + \frac{C}{2}\sum_{i=1}^{l}\xi_i^2 \\ &\text{subject to } \ y_i = \mathbf{w}\cdot\phi(\mathbf{x}_i) + b + \xi_i \end{aligned} \tag{3}$$

and the corresponding primal Lagrangian gives the following unconstrained minimization problem:

$$\mathcal{L} = \frac{1}{2}\|\mathbf{w}-\beta\mathbf{w}'\|^2+\frac{C}{2}\sum_{i=1}^{l}\xi_i^2-\sum_{i=1}^{l}\alpha_i\{\mathbf{w}\cdot\phi(\mathbf{x}_i)+b+\xi_i-y_i\}\ , \tag{4}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_l) \in \mathbb{R}^l$ is the vector of Lagrange multipliers. The optimality conditions can be expressed as:

$$\frac{\partial\mathcal{L}}{\partial\mathbf{w}} = \mathbf{0} \implies \mathbf{w} = \beta\mathbf{w}' + \sum_{i=1}^{l}\alpha_i\phi(\mathbf{x}_i)\ , \tag{5}$$

$$\frac{\partial\mathcal{L}}{\partial b} = 0 \implies \sum_{i=1}^{l}\alpha_i = 0\ , \tag{6}$$

$$\frac{\partial\mathcal{L}}{\partial\xi_i} = 0 \implies \alpha_i = C\xi_i\ , \tag{7}$$

$$\frac{\partial\mathcal{L}}{\partial\alpha_i} = 0 \implies \mathbf{w}\cdot\phi(\mathbf{x}_i) + b + \xi_i - y_i = 0\ . \tag{8}$$

From (5) it is clear that the adapted model is given by the sum of the pre-trained model $\mathbf{w}'$ (weighted by β) and a new model $\mathbf{w}$ obtained from the new samples. Note that when β is 0 we recover the original LS-SVM formulation without any adaptation to previous data. Using (5) and (7) to eliminate $\mathbf{w}$ and $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_l)$ from (8) we find that:

$$\sum_{j=1}^{l}\alpha_j\ \phi(\mathbf{x}_j)\cdot\phi(\mathbf{x}_i) + b + \frac{\alpha_i}{C} = y_i - \beta\mathbf{w}'\cdot\phi(\mathbf{x}_i)\ . \tag{9}$$

Denoting with $\boldsymbol{K}$ the kernel matrix, i.e. $\boldsymbol{K}_{ji} = \boldsymbol{K}(\mathbf{x}_j, \mathbf{x}_i) = \phi(\mathbf{x}_j)\cdot\phi(\mathbf{x}_i)$, the obtained system of linear equations can be written more concisely in matrix form as:

$$\begin{bmatrix} \mathbf{K} + \frac{1}{C}\mathbf{I} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{\alpha} \\ b \end{bmatrix} = \begin{bmatrix} \mathbf{y} - \beta\mathbf{w}'\cdot\phi(\mathbf{x}) \\ 0 \end{bmatrix}\ . \tag{10}$$

Thus the model parameters can be calculated with:

$$\begin{bmatrix} \boldsymbol{\alpha} \\ b \end{bmatrix} = \boldsymbol{M}^{-1}(\boldsymbol{Y} - \beta\hat{\boldsymbol{Y}}) \tag{11}$$

where $\boldsymbol{Y}$ is the vector $[y_1, \ldots, y_l, 0]^T$, $\hat{\boldsymbol{Y}}$ is the vector containing the predictions of the previous model $[\mathbf{w}'\cdot\phi(\mathbf{x}_1), \ldots, \mathbf{w}'\cdot$

$\phi(\mathbf{x}_l), 0]^T$, and $\boldsymbol{M}$ is the first matrix on the left in (10). Let $[\boldsymbol{\alpha}', b']^T = \boldsymbol{M}^{-1}\boldsymbol{Y}$ and $[\boldsymbol{\alpha}'', b'']^T = \boldsymbol{M}^{-1}\hat{\boldsymbol{Y}}$, from the equation above and using the same steps in [27] (see the Appendix), we have the following

Proposition 1. *The prediction on sample i, when removed from the training set, is [10]:*

$$\tilde{y}_i = y_i - \frac{\alpha'_i}{\boldsymbol{M}^{-1}_{ii}} + \beta \frac{\alpha''_i}{\boldsymbol{M}^{-1}_{ii}} \ . \tag{12}$$

From (12) the leave-one-out-error is easily evaluated, according to the required measure of accuracy for the problem at hand.

Notice that, in the above formula, β is the only parameter, hence, it is possible to set it optimally in order to minimize the leave-one-out error, while at the same time choosing the best pre-trained model for adaptation. Moreover, $\boldsymbol{\alpha}$ depends linearly on β, thus it is easy to define the learning model which is fixed once β has been chosen.

The complexity of the algorithm is dominated by the evaluation of the matrix $\boldsymbol{M}$, which must anyway occur while training; thus, the computational complexity of evaluating the leave-one-out error is negligible, if compared to the complexity of training. As a last remark, we underline that the pre-trained model $\mathbf{w}'$ can be obtained by any training algorithm, as far as it can be expressed as a weighted sum of kernel functions. The framework is therefore very general.

C. Model Adaptation from Multiple Subjects

The approach described in the previous Section has a main drawback: although many prior knowledge models are available, it uses only one of them, selected as the most useful in term of minimal leave-one-out error. Even if the pre-trained models are not equally informative, relying on more than one of them may be beneficial. To this goal it is possible to define a new learning problem which considers the linear combination of all the known models [29]:

$$\begin{aligned} \min_{\mathbf{w},b} \ & \frac{1}{2} \left\| \mathbf{w} - \sum_{j=1}^{N} \beta_j \mathbf{w}'_{\mathbf{j}} \right\|^2 + \frac{C}{2} \sum_{i=1}^{l} \xi_i^2 \\ \text{subject to} \quad & y_i = \mathbf{w} \cdot \phi(\mathbf{x}_i) + b + \xi_i \ . \end{aligned} \tag{13}$$

The original single coefficient β has been substituted with a vector $\boldsymbol{\beta}$ containing as many elements as the number of prior models, N. For this formulation the optimal solution is:

$$\mathbf{w} = \sum_{j=1}^{N} \beta_j \mathbf{w}'_{\mathbf{j}} + \sum_{i=1}^{l} \alpha_i \phi(\mathbf{x}_i) \ . \tag{14}$$

Here $\mathbf{w}$ is expressed as a weighted sum of the pre-trained models scaled by the parameters β_j, plus the new model built on the incoming training data [29]. The leave-one-out prediction of each sample i can again be written in closed form as:

$$\tilde{y}_i = y_i - \frac{\alpha'_i}{\boldsymbol{M}^{-1}_{ii}} + \sum_{j=1}^{N} \beta_j \frac{\alpha''_{i(j)}}{\boldsymbol{M}^{-1}_{ii}} \ . \tag{15}$$

Now $[\boldsymbol{\alpha}''_{i(j)}, b''_j]^T = \boldsymbol{M}^{-1}\hat{\mathbf{Y}}_{\mathbf{j}}$ and $\hat{\mathbf{Y}}_{\mathbf{j}}$ is the vector which contains the predictions of the j^{th} previous model $[\mathbf{w}'_j \cdot \phi(\mathbf{x}_1), \ldots, \mathbf{w}'_j \cdot \phi(\mathbf{x}_l), 0]$. As before, the leave-one-out error can be calculated and minimized to evaluate the best weights β_j.

III. Learning How Much to Adapt

The adaptive learning methods described above look for the model parameters ($\mathbf{w}$,b) once the value of the weight β, or the corresponding vector $\boldsymbol{\beta}$, has been chosen. This choice actually defines another learning problem that gives as output an indication of how much each of the pre-trained models are reliable for adaptation. In the following we define how to face this issue in the classification and regression cases, a general scheme of the proposed solutions is in Figure 2.

A. Classification

For a binary classification problem, and in case of a single pre-trained model, we can follow the approach proposed in [27] and find β by minimizing the continuous and derivable logistic function :

$$\min_{\beta} \sum_{i=1}^{l} \frac{1}{1 + \exp(-10 e_i)} \ , \tag{16}$$

based on the leave-one-out error $e_i = \tilde{y}_i - y_i$. However in our application we have multiple pre-trained models and multiple classes $y_i \in \{1, \ldots, G\}$ where G is the number of analyzed grasp types. We can directly extend the methods presented in the previous Section using the one-vs-all approach. In this way we obtain G different leave-one-out predictions $\tilde{y}_{i(y)}$ for each sample i. It is necessary to compose all these values to define a single multiclass estimate of the leave-one-out error.

1) Best Prior Model: A first solution could be to consider:

$$e_i = \max_{y \neq y_i} \mid \tilde{y}_{i(y_i)} - \tilde{y}_{i(y)} \mid \ , \tag{17}$$

and to evaluate (16) for each of the $j \in \{1, \ldots, N\}$ pre-trained models, varying β with small steps in $[0, 1]$ (this is the approach used in [10]). The minimal result identifies both the best known subject for adaptation and, at the same time, the corresponding β weight. Still, this approach is non-convex thus the existence of a global optimum is not guaranteed.

2) Multiple Prior Models: When moving to the choice of multiple weights for all the pre-trained models we can also overcome the non-convexity issue described above, by minimizing the loss function proposed in [29]:

$$L(y_i, \tilde{y}_i) = \max \ [\ 1 - y_i \tilde{y}_i \ , \ 0 \ ] \ , \tag{18}$$

this is a convex upper-bound to the leave-one-out misclassification loss and has a smoothing effect, similar to the logistic function in (16). However this choice is again suitable only for binary problems. Here we propose to use the convex multiclass loss [30]:

$$L(y_i, \tilde{y}_i) = \max_{y \neq y_i} \ [\ 1 \ + \ \tilde{y}_{i(y)} - \tilde{y}_{i(y_i)} \ , 0] \ , \tag{19}$$

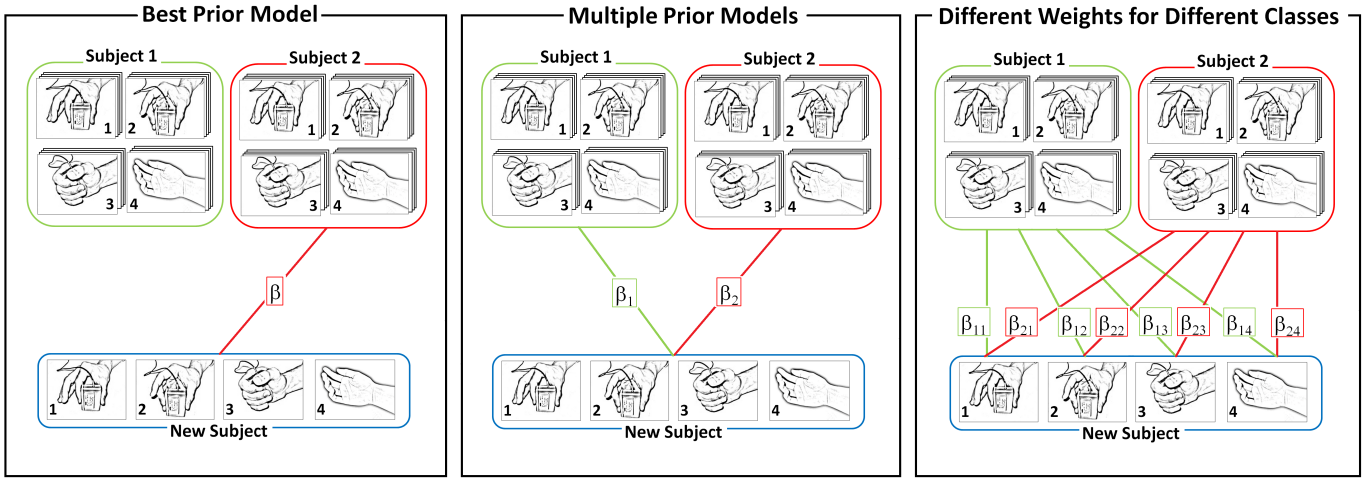

Fig. 2. This figure shows the three methods adopted to leverage information from multiple known subjects when learning on a new one. For all the known subjects many sEMG signal samples are available, while few sEMG signals are recorded from the new subject. Left: choose only the best known subject and use its reweighted model as a starting point for learning. Center: consider a linear combination of the known subjects with equal weight for all the grasp models of each subject. Right: consider again a linear combination of all the known models but assign a different weight to each grasp model for each subject.

with the final objective function:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{l} L(y_i, \tilde{y}_i) \quad \text{subject to} \quad \|\boldsymbol{\beta}\|_2 \leq 1 \ . \tag{20}$$

The condition of having $\boldsymbol{\beta}$ in the unitary ball is similar to the regularization term used in LS-SVM and it is a natural generalization of the original constraint $\beta \in [0, 1]$. This term is necessary to avoid overfitting problems which can happen when the number of known models is large compared to the number of training samples [29].

We implemented the optimization process using a simple projected sub-gradient descendent algorithm, where at each iteration $\boldsymbol{\beta}$ is projected onto the l_2-sphere, $\|\boldsymbol{\beta}\|_2 \leq 1$. The pseudo-code is in Algorithm 1.

3) Different Weights for Different Classes: Till now we considered techniques which assign a unique weight to each known subject. This means that, the whole set of one-vs-all pre-trained models for a subject are equally weighted. However, when learning the model for class $y = 1$, it may be useful to give more weight in adaptation to subject $j = 1$ than $j = 2$, while it could be the opposite when learning the model for class $y = 2$. To have one more degree of freedom and decide the adaptation specifically for each class, we enlarge the set of weight parameters introducing the matrix $\mathbf{B} \in \mathbb{R}^{N \times G}$ where each row $j \in \{1, \ldots, N\}$ contains the vector $\boldsymbol{\beta}_j^T$ with G elements, one for each class.

The optimization problem is analogous to the one described in (20), with a changing in the constraints. Each class problem is now considered separately, so we have G conditions, each for one of the columns of the $\mathbf{B}$ matrix: naming $\mathbf{B}_{*,y} = [B_{1,y}, B_{2,y}, \ldots B_{N,y}]^T$ for a fixed $y \in \{1, \ldots, G\}$, a single constraint is $\|\mathbf{B}_{*,y}\|_2 \leq 1$.

Algorithm 1 Projected Sub-gradient Descent Algorithm

Notice that at line 7, we slightly abuse the max notation, meaning that each element of the $\boldsymbol{\beta}^k$ vector is changed to zero if negative.

Input: $\mathbf{L}(\boldsymbol{\beta}) = [L(y_1, \tilde{y}_1), L(y_2, \tilde{y}_2), \ldots, L(y_l, \tilde{y}_l)]^T$

1: $\boldsymbol{\beta}^0 = \mathbf{0} \in \mathbb{R}^N$
2: $k = 1$
3: **repeat**
4: $\quad \frac{\partial L}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}^{k-1}) = \mathbf{D} \in \mathbb{R}^{N \times l}$
5: $\quad \mathbf{d} \in \mathbb{R}^N, \ \ \mathbf{d}_j = \sum_{i=1}^{l} \mathbf{D}_{ji}$
6: $\quad \boldsymbol{\beta}^k = \boldsymbol{\beta}^{k-1} - \frac{1}{\sqrt{k}} \mathbf{d}$
7: $\quad \boldsymbol{\beta}^k = \max[\boldsymbol{\beta}^k, \mathbf{0}]$
8: $\quad$ **if** $\|\boldsymbol{\beta}^k\|_2 > 1$ **then**
9: $\quad\quad \boldsymbol{\beta}^k = \boldsymbol{\beta}^k / \|\boldsymbol{\beta}^k\|_2$
10: $\quad$ **end if**
11: $\quad k = k + 1$
12: **until convergence**

Output: $\boldsymbol{\beta}$

B. Regression

Our goal in using regression is the prediction of the force applied by one subject in grasping, independently to the specific kind of grasp performed. Thus now the output y_i for each corresponding input $\mathbf{x}_i$ is a continuous real value, rather a discrete one as in classification.

Similarly to what seen before, it is possible to learn the regression model relying on information from the closest known subject, or on the combination of multiple pre-trained models.

1) Best Prior Model: We can use the leave-one-out prediction in (12) to evaluate the Mean Square Error (MSE):

$$\text{MSE} = \frac{1}{l}\sum_{i=1}^{l}(y_i - \tilde{y}_i)^2 = \frac{1}{l}\sum_{i=1}^{l}\left(\frac{\alpha'_i}{\boldsymbol{M}_{ii}^{-1}} + \beta\frac{\alpha''_i}{\boldsymbol{M}_{ii}^{-1}}\right)^2 . \quad (21)$$

This is a quadratic function in β and the minimum is obtained using:

$$\beta = \frac{\left[\sum_{i=1}^{l}\frac{\alpha'_i}{\boldsymbol{M}_{ii}^{-1}}\frac{\alpha''_i}{\boldsymbol{M}_{ii}^{-1}}\right]}{\left[\sum_{i=1}^{l}\left(\frac{\alpha''_i}{\boldsymbol{M}_{ii}^{-1}}\right)^2\right]} . \quad (22)$$

We use the constraint $\beta \geq 0$, just imposing $\beta = 0$ every time it results negative. Once calculated the minimum MSE value for each $j \in \{1, 2, \ldots N\}$, comparing all of them, we can identify the best known subject to use for adaptation when learning the regression model on a new subject.

2) Multiple Prior Models: To take advantage from all the available pre-trained models we can combine them linearly and search for a vector of weights as in classification. In this way we obtain:

$$\text{MSE} = \frac{1}{l}\sum_{i=1}^{l}\left(\frac{\alpha'_i}{\boldsymbol{M}_{ii}^{-1}} + \sum_{j=1}^{N}\beta_j\frac{\alpha''_{i(j)}}{\boldsymbol{M}_{ii}^{-1}}\right)^2 . \quad (23)$$

Adding also the condition $\|\boldsymbol{\beta}\|_2 \leq 1$, we can find the best $\boldsymbol{\beta}$ vector which minimizes the MSE with a Quadratically Constrained Quadratic Program (QCQP) solver. We used CVX [31] in our experiments.

IV. Experimental Data

To test the effectiveness of our model adaptation techniques we use the database of sEMG / hand posture / force signals already presented and described in [9] and already used in [9], [10]. (The following description of the database is very concise; the interested reader should refer to the above cited paper(s) for more details.) Data is collected from 10 intact subjects (2 women, 8 men) using 7 sEMG electrodes (Aurion ZeroWire wireless) placed on the dominant forearm according to the medical literature [32]. A FUTEK LMD500 force sensor [33] is used to measure the force applied by the subject's hand during the recording. Data is originally sampled at 2kHz. Each subject starts from a rest condition (sEMG baseline activity) then repeatedly grasps the force sensor using, in turn, three different grips, visible in Figure 3. The subject either remains seated and relaxed while performing the grasps, or is free to move (walk around, sit down stand up, etc.). These phases are referred to as *Still-Arm (SA)* and *Free-Arm (FA)* respectively. Each grasping action is repeated along 100 seconds of activity. The whole procedure is repeated twice. The root mean square of the signals along 1s (for classification) and 0.2s (for regression) is evaluated; subsampling at 25Hz follows. Samples for which the applied force is lower than 20% of the mean force value obtained for each subject are labeled as "rest" class. After this pre-processing we got around 15000 samples per subject, each sample consists of a 7 elements sEMG signal vector and one force value.

V. Experimental Results

As already mentioned in Section II-B, our working assumption is to have N pre-trained models stored in memory; new data comes from subject $N+1$ and the system starts training, to build the $N+1$'th model. The performance is then evaluated using unseen data from subject $N+1$. To simulate this scenario and to have a reliable estimation of the performance, we use a leave-one-out approach: out of the 10 subjects for which we have data recordings, we train 9 models off-line. These correspond to the N stored models in memory, while data from the remaining subject are used for the adaptive learning of the $N+1$'th model. This procedure is repeated 10 times, using in turn all the recorded subjects for the adaptive learning of the model.

The training sequences are random subsets from the entire dataset of the new subject, i.e. they are taken without considering the order in which they were acquired. We considered 12 successive learning steps, for each of them the number of available training samples increases by 30 elements reaching a maximum of 360 samples. The test runs over all the remaining samples.

We conducted two sets of experiments to analyze the behavior of the proposed adaptive learning methods in different conditions: we used pre-trained models evaluated on SA data, while for each new subject we considered (1) data recorded in SA condition and (2) data recorded in FA condition. We both classify the grasp type and predict the force measured by the force sensor.

We name the proposed adaptation models respectively:

- *Best-Adapt*: adaptive learning starting from the best prior knowledge model [10];
- *Multi-Adapt*: adaptive learning starting from a linear combination of the known models;
- *Multi-perclass-Adapt*: adaptive learning (for classification) starting from a linear combination of the known models with a different weight for each class.

To assess the performance of all these methods we compare them to two baseline approaches:

- *Prior*: consists in using only the pre-trained models without updating them with the new training data. We consider their average performance.
- *No-Adapt*: is plain LS-SVM using only the new data for training, as it would be in the standard scenario without adaption.

As a measure of performance, for classification we use the standard classification rate; for regression, the performance index is the correlation coefficient evaluated between the predicted force signal and the real one. Although we minimized the Mean Square Error in the regression learning process, the choice of the correlation coefficient is suggested by a practical consideration: when driving a prosthesis, or even a non-prosthetic mechanical hand, we are not interested in the absolute force values desired by the user / subject: mechanical

(a) (b) (c)

Fig. 3. The three different grips employed in the experiment: (a) index precision grip; (b) other fingers precision grip; (c) power grasp. Reproduced from [9].

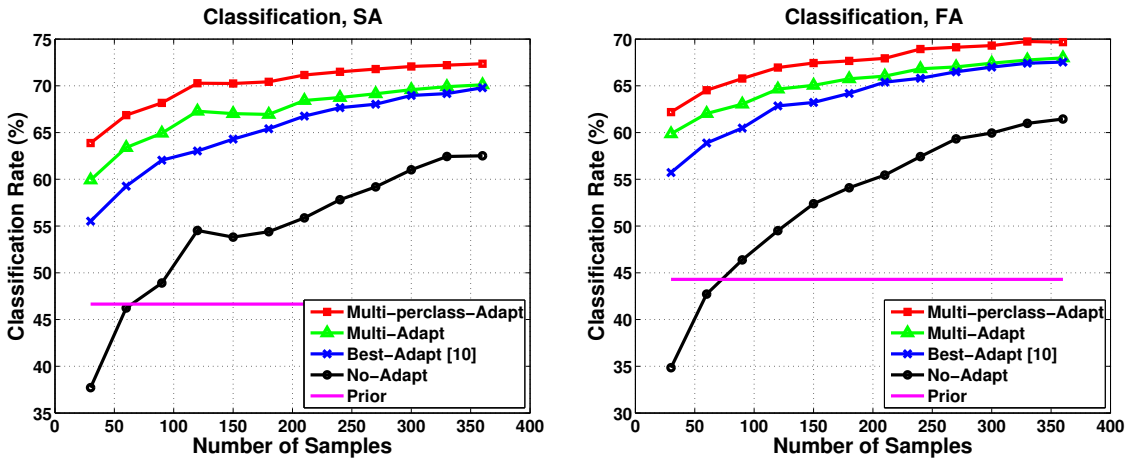

Fig. 4. Classification rate obtained averaging over all the subjects as a function of the number of samples in the training set. The title of each figure specifies if the data used are registered in Still-Arm (SA) or Free-Arm (FA) setting.

hands usually cannot apply as much force as human hands do, for obvious safety reasons, or e.g., in teleoperation scenarios, they could be able to apply *much more* force than a human hand can. As already done, e.g. in [11], [13], [9], we are rather concerned with getting a signal which is *strongly correlated* with the user / subject's will. The significance of the comparisons between the methods is evaluated through the sign test [34].

To build the pre-trained models we used the standard SVM algorithm. All the parameters to be set during training (C and γ of the Gaussian kernel) were chosen by cross-validation.

Figure 4(left) reports the obtained classification rate at each step when using SA data. The plot shows that Multi-perclass-Adapt outperforms both the baselines No-Adapt, Prior, and all the other adaptive learning methods. The gain obtained by Multi-perclass-Adapt with respect to No-Adapt ($p < 0.003$) stabilizes around 10% for 300-360 training samples. The difference between Multi-perclass-Adapt and Best-Adapt shows an average advantage in recognition rate of around 2.5% ($p < 0.03$). Analogous results are obtained when considering FA data for the new subject. In this case we are trying to reproduce a more realistic scenario where the prior knowledge is built on data recorded on 9 subjects in laboratory controlled conditions (Still-Arm) while the new subject moves freely. Figure 4(right) reports the classification rate results in this setting. Multi-perclass-Adapt shows again the best performance, with around 2% recognition gain with respect to Best-Adapt ($p < 0.03$). A significant difference it is also evident between Multi-perclass-Adapt and No-Adapt ($p < 0.03$) and reaches 8% in recognition rate for 360 samples.

Analyzing Figure 4 as a whole, we can state that all the proposed adaptive methods outperform learning from scratch with the best results obtained when exploiting a linear combination of pre-trained models with a different weight for each known subject and each class (Multi-perclass-Adapt). Moreover, we notice that learning with adaptation with 30 training samples performs almost as No-Adapt with around 300 samples. Considering the acquisition time, this means that the adaptive methods are almost ten time faster than learning from scratch. Using the prior knowledge by itself appears as a good choice if only very few training samples are available but looses its advantage when the dimension of the training set increases. Passing from SA to FA data we can also notice that the results for Prior show a small drop (from 46.7% to 44.3%) related to the change in domain between the data used for pre-trained model and the one used for the new subject.

The corresponding regression results are reported in Figure 5. From the plot on the left we can notice that both the adap-

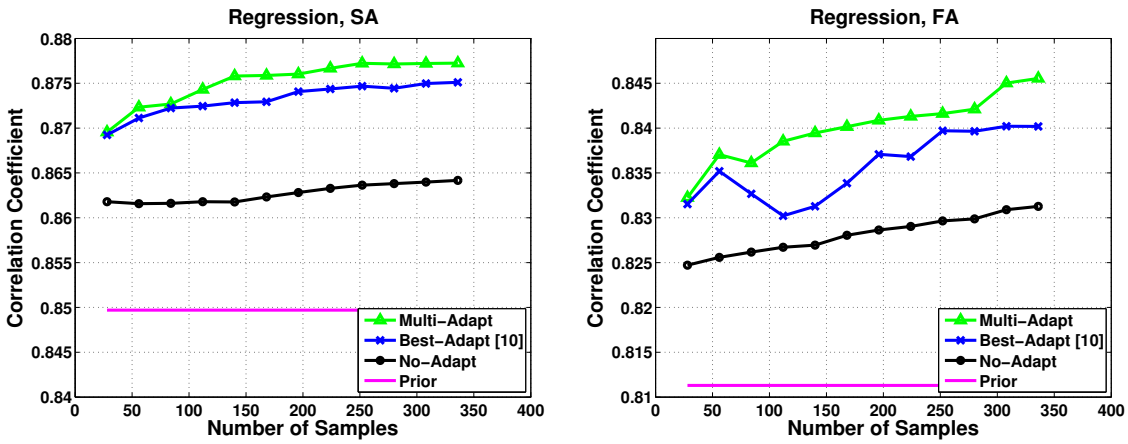

Fig. 5. Correlation coefficient obtained averaging over all the subjects as a function of the number of samples in the training set. The title of each figure specifies if the data used are registered in Still-Arm (SA) or Free-Arm (FA) setting.

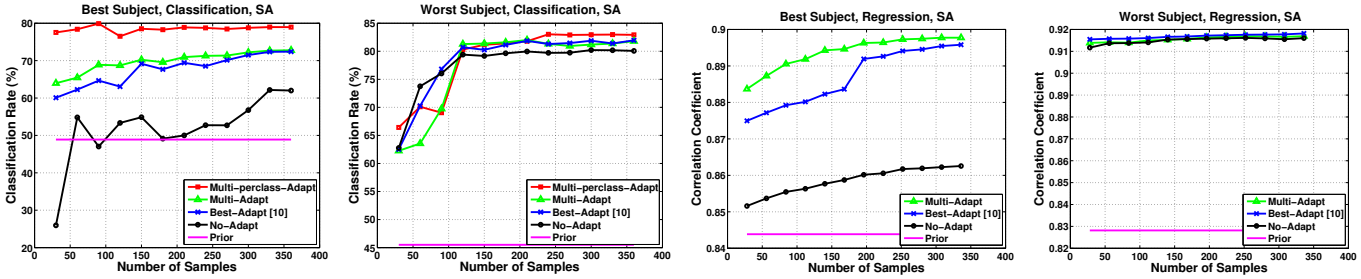

Fig. 6. Classification and Regression over SA data for the best and worst subjects. With best and worst we mean the subjects for which the difference in performance between learning with adaptation and learning from scratch is respectively maximum and minimum.

tive learning methods highly outperform No-Adapt and Prior ($p < 0.003$). However in this case Multi-Adapt and Best-Adapt performs almost equally (no statistical significant difference). Figure 5(right) shows that Best-Adapt has a less robust and less smooth performance with respect to Multi-Adapt when passing to the setting based on FA data in training. Still the two methods are statistically equivalent and they show a significant gain with respect to No-Adapt only for more than 250 training samples ($p < 0.03$). The higher difficulty in the FA task with respect to the SA is demonstrated, as for classification, by the drop in Prior performance (from 0.849 to 0.811). Although we decided to show the correlation coefficient results, the corresponding Mean Square Error would lead to the same conclusions.

Discussion

As a general remark we can state that the three proposed adaptive methods (Multi-perclass-adapt, Multi-Adapt and Best-Adapt), improve the learning performance to different extents if prior-knowledge contains useful information for the new task, and never harm if any good match between the data of the new subject and the old source subjects is found. To further demonstrate this statement, Figure 6 shows the classification and regression results on SA data respectively for the subject that have the maximum (best) and the minimum (worst) difference in recognition and regression performance with adaptation compared to No-Adapt. The worse-case subject represents the paradigmatic case of no previous models matching the current distribution; as a consequence the parameter β ($\boldsymbol{\beta}$) is set automatically to a small value (to a vector of small norm). In this case there is essentially no transfer of prior knowledge. However it is reasonable to claim that the overall performance of the method would increase along with the number of stored models, since this would mean a larger probability of finding matching pre-trained models. In the long run, a large database of sEMG signals and force measure, with subjects possibly categorized (per age, sex, body characteristics, etc.) in order to avoid too hard a computational burden, would definitively help getting uniformly better performance.

Regarding the use of prior knowledge on a new problem, we point out here that, without an appropriate way to choose how to weight and combine it with the new information, it is only partially useful. In fact, the Prior line in all the plots for our results corresponds to an attempt to use directly a flat combination of all the pre-trained models on a new subject: the obtained performance show that this is not a good solution.

Finally we briefly discuss the choice of the learning parameter C. Here we followed the standard approach in the community, and kept the the parameter C fixed using the best value obtained from cross validation on the known subjects. Still, one might argue that the best way to define it is to optimize it by using the available training samples of the target subject, separately for each learning approach. For the

proposed adaptive methods, this would imply to define C together with β, leading to a non-convex problem with also a great increase in computational complexity.

VI. Conclusion

The results presented in this paper clearly show that machine-learning-based classification and regression applied to surface EMG can be improved by means of re-using previous knowledge. In particular, we have used optimal SVM models previously built by training on a pool of human subjects to boost the training time and adaptation of a LS-SVM to new subjects. All the proposed adaptive methods, when the learning is performed on 300 training samples show a gain of around 8-10% in recognition rate for grasp type classification and 0.014 in correlation coefficient for regression when predicting the applied force, with respect to learning from scratch on the new subject. These results have been obtained on a database of sEMG / hand posture / force data already used in [9], [10], consisting of data collected from 10 intact subjects in controlled and non-controlled conditions. In particular, in the non-controlled condition the subjects were walking, sitting and standing and moving the arms freely while recording.

The overall idea is that a prosthesis could be embedded with this additional, pre-existing knowledge before being shipped out to the generic patient. This needs to be done once and for all and, most likely, for a large pool of healthy subjects and/or amputees of diverse condition, age and type of operation, and degree of muscle remnant fitness. The fact that the free-arm condition consistently benefits as well from the proposed technique — essentially to the same extent as the controlled one — is a very promising hint, meaning that one could potentially pre-train a prosthesis *in a laboratory* and then ship it outside, and still give a significant benefit of the patient with respect to the learning-from-scratch case.

The database we used consists of intact subjects only, but it is believed that trans-radial amputees can generate similarly accurate signals ([14] is the most recent result on this topic), so this seems no disruptive objection to the applicability of the method. The project NinaPro (http://www.idiap.ch/project/ninapro/) is currently concerned with collecting such a large database of mixed subjects. If confirmed in the large scale, the current result could pave the way to a significantly higher acceptance of myoprostheses in the clinical setting.

Acknowledgment

This work is partially supported by the Swiss National Science Foundation *Sinergia* project Ninapro (Non-Invasive Adaptive Prosthetics). T. T. was supported by the EMMA Hasler Project. We are thankful to Arjan Gijsberts for his help in cleaning the code for the experiments.

The database used in this work has been collected and refined in 2009 at the Robotics, Brain and Cognitive Sciences Department of the Italian Institute of Technology, Genova, Italy, mainly by Angelo Emanuele Fiorilla. We would like to thank him and Giulio Sandini of the same Institute for making the database available for this work.

Appendix
Closed formula for the leave-one-out prediction

We show here that, following the same steps presented in [27], it is possible to demonstrate the Proposition 1 obtaining the closed formula for the leave-one-out prediction in (12). We start from

$$\boldsymbol{M}\begin{bmatrix}\boldsymbol{\alpha}\\ b\end{bmatrix} = \begin{bmatrix}\mathbf{y} - \beta\mathbf{w}'\cdot\phi(\mathbf{x})\\ 0\end{bmatrix} . \tag{24}$$

and we decompose $\boldsymbol{M}$ into block representation isolating the first row and column as follows:

$$\boldsymbol{M} = \begin{bmatrix}\mathbf{K}+\frac{1}{C}\mathbf{I} & \mathbf{1}\\ \mathbf{1}^T & 0\end{bmatrix} = \begin{bmatrix}m_{11} & \mathbf{m}_1^T\\ \mathbf{m}_1 & \mathbf{M}_1\end{bmatrix} . \tag{25}$$

Let $[\boldsymbol{\alpha}^{(-i)}, b^{(-i)}]^T$ represent the parameter of LS-SVM during the i^{th} iteration of the leave-one-out cross validation procedure. In the first iteration, where the first training sample is excluded we have

$$\begin{bmatrix}\boldsymbol{\alpha}^{(-1)}\\ b^{(-1)}\end{bmatrix} = \mathbf{M}_1^{-1}(\mathbf{Y}_1 - \beta\hat{\mathbf{Y}}_1) , \tag{26}$$

where $\mathbf{Y}_1 = [y_2, \ldots, y_l, 0]^T$ and $\hat{\mathbf{Y}}_1 = [\mathbf{w}'\cdot\phi(\mathbf{x}_2), \ldots, \mathbf{w}'\cdot\phi(\mathbf{x}_l), 0]^T$. The leave-one-out prediction for the first training sample is then given by

$$\begin{aligned}\tilde{y}_1 &= \mathbf{m}_1^T\begin{bmatrix}\boldsymbol{\alpha}^{(-1)}\\ b^{(-1)}\end{bmatrix} + \beta\mathbf{w}'\cdot\phi(\mathbf{x}_1) &(27)\\ &= \mathbf{m}_1^T\mathbf{M}_1^{-1}(\mathbf{Y}_1 - \beta\hat{\mathbf{Y}}_1) + \beta\mathbf{w}'\cdot\phi(\mathbf{x}_1) . &(28)\end{aligned}$$

Considering the last l equations in the system in (24), it is clear that $[\mathbf{m}_1\ \mathbf{M}_1][\alpha_1, \ldots, \alpha_l, b]^T = (\mathbf{Y}_1 - \beta\hat{\mathbf{Y}}_1)$, and so

$$\begin{aligned}\tilde{y}_1 &= \mathbf{m}_1^T\mathbf{M}_1^{-1}[\mathbf{m}_1\mathbf{M}_1][\boldsymbol{\alpha}, b]^T + \beta\mathbf{w}'\cdot\phi(\mathbf{x}_1)\\ &= \mathbf{m}_1^T\mathbf{M}_1^{-1}\mathbf{m}_1\alpha_1 + \mathbf{m}_1^T[\alpha_2, \ldots, \alpha_l, b]^T + \beta\mathbf{w}'\cdot\phi(\mathbf{x}_1) .\end{aligned} \tag{29}$$

Noting from the first equation in the system in (24) that $y_1 - \beta\mathbf{w}'\cdot\phi(\mathbf{x}_1) = m_{11}\alpha_1 + \mathbf{m}_1^T[\alpha_2, \ldots, \alpha_l, b]^T$, we have

$$\tilde{y}_1 = y_1 - \alpha_1(m_{11} - \mathbf{m}_1^T\mathbf{M}_1^{-1}\mathbf{m}_1) . \tag{30}$$

Finally, via the block matrix inversion lemma,

$$\mathbf{M}^{-1} = \begin{bmatrix}\mu^{-1} & -\mu^{-1}\mathbf{m}_1\mathbf{M}_1^{-1}\\ \mathbf{M}_1^{-1} + \mu^{-1}\mathbf{M}_1^{-1}\mathbf{m}_1^T\mathbf{m}_1\mathbf{M}_1^{-1} & -\mu^{-1}\mathbf{M}_1^{-1}\mathbf{m}_1^T\end{bmatrix} ,$$

where $\mu = m_{11} - \mathbf{m}_1^T\mathbf{M}_1^{-1}\mathbf{m}_1$ and noting that the system of linear equations (24) is insensitive to permutations of the ordering of the equations and of the unknowns, we have

$$\tilde{y}_i = y_i - \frac{\alpha_i}{\mathbf{M}_{ii}^{-1}} . \tag{31}$$

Let $[\boldsymbol{\alpha}', b']^T = \boldsymbol{M}^{-1}\boldsymbol{Y}$ and $[\boldsymbol{\alpha}'', b'']^T = \boldsymbol{M}^{-1}\hat{\boldsymbol{Y}}$, where now $\mathbf{Y} = [y_1, \ldots, y_l, 0]^T$ and $\hat{\mathbf{Y}} = [\mathbf{w}'\cdot\phi(\mathbf{x}_1), \ldots, \mathbf{w}'\cdot\phi(\mathbf{x}_l), 0]^T$, from the equation above we get :

$$\tilde{y}_i = y_i - \frac{\alpha_i'}{\mathbf{M}_{ii}^{-1}} + \beta\frac{\alpha_i''}{\mathbf{M}_{ii}^{-1}} . \tag{32}$$

REFERENCES

[1] M. Zecca, S. Micera, M. C. Carrozza, and P. Dario, "Control of multifunctional prosthetic hands by processing the electromyographic signal," *Critical Reviews in Biomedical Engineering*, vol. 30, no. 4–6, pp. 459–485, 2002.

[2] S. Micera, J. Carpaneto, and S. Raspopovic, "Control of hand prostheses using peripheral information," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 48–68, Oct. 2010.

[3] B. Peerdeman, D. Boere, H. Witteveen, R. Huis in 't Veld, H. Hermens, S. Stramigioli, H. Rietman, P. Veltink, and S. Misra, "Myoelectric forearm prostheses: State of the art from a user-centered perspective," *J Rehabil Res Dev.*, vol. 48, no. 6, pp. 719–738, 2011.

[4] D. J. Atkins, D. C. Y. Heard, and W. H. Donovan, "Epidemiologic overview of individuals with upper-limb loss and their reported research priorities," *Journal Of Prosthetics And Orthotics*, vol. 8, no. 1, pp. 2–11, 1996.

[5] A. H. Bottomley, "Myoelectric control of powered prostheses," *J Bone Joint Surg*, vol. B47, pp. 411—415, 1965.

[6] D. S. Childress, "A myoelectric three-state controller using rate sensitivity," in *Proceedings 8th ICMBE, Chicago, IL*, 1969, pp. 4—5.

[7] L. Philipson, D. S. Childress, and J. Strysik, "Digital approaches to myoelectric state control of prostheses," *Bulletin of Prosthetics Research*, vol. 18, no. 2, pp. 3—11, 1981.

[8] F. C. P. Sebelius, B. N. Rosén, and G. N. Lundborg, "Refined myoelectric control in below-elbow amputees using artificial neural networks and a data glove," *Journal of Hand Surgery*, vol. 30A, no. 4, pp. 780—789, 2005.

[9] C. Castellini, A. E. Fiorilla, and G. Sandini, "Multi-subject / daily-life activity EMG-based control of mechanical hands," *Journal of Neuroengineering and Rehabilitation*, vol. 6, no. 41, 2009.

[10] F. Orabona, C. Castellini, B. Caputo, E. Fiorilla, and G. Sandini, "Model adaptation with least-squares SVM for hand prosthetics," in *Proceedings of ICRA - International Conference on Robotics and Automation*, 2009, pp. 2897–2903.

[11] C. Castellini and P. van der Smagt, "Surface EMG in advanced hand prosthetics," *Biological Cybernetics*, vol. 100, no. 1, pp. 35–47, 2009.

[12] F. V. Tenore, A. Ramos, A. Fahmy, S. Acharya, R. Etienne-Cummings, and N. V. Thakor, "Decoding of individuated finger movements using surface electromyography," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1427—1434, 2009.

[13] C. Castellini, E. Gruppioni, A. Davalli, and G. Sandini, "Fine detection of grasp force and posture by amputees via surface electromyography," *Journal of Physiology (Paris)*, vol. 103, no. 3-5, pp. 255–262, 2009.

[14] C. Cipriani, C. Antfolk, M. Controzzi, G. Lundborg, B. Rosen, M. Carrozza, and F. Sebelius, "Online myoelectric control of a dexterous hand prosthesis by transradial amputees," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 19, no. 3, pp. 260–270, june 2011.

[15] R. Merletti, M. Aventaggiato, A. Botter, A. Holobar, H. Marateb, and T. Vieira, "Advances in surface EMG: Recent progress in detection and processing techniques," *Critical reviews in biomedical engineering*, vol. 38, no. 4, pp. 305—345, 2011.

[16] R. Merletti, A. Botter, C. Cescon, M. Minetto, and T. Vieira, "Advances in surface EMG: Recent progress in clinical research applications," *Critical reviews in biomedical engineering*, vol. 38, no. 4, pp. 347—379, 2011.

[17] R. Merletti, A. Botter, A. Troiano, E. Merlo, and M. Minetto, "Technology and instrumentation for detection and conditioning of the surface electromyographic signal: State of the art," *Clinical Biomechanics*, vol. 24, pp. 122—134, 2009.

[18] T. Lorrain, N. Jiang, and D. Farina, "Influence of the training set on the accuracy of surface EMG classification in dynamic contractions for the control of multifunction prostheses," *J Neuroeng Rehabil*, vol. 8:25, 2011.

[19] Q. Sun, R. Chattopadhyay, S. Panchanathan, and J. Ye, "A two-stage weighting framework for multi-source domain adaptation," in *Advances in Neural Information Processing Systems (NIPS)*, 2011.

[20] H. Daumé III and D. Marcu, "Domain adaptation for statistical classifiers," *Journal of Artificial Intelligence Research (JAIR)*, vol. 26, pp. 101–126, 2006.

[21] S. B. David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine Learning*, vol. 79, no. 1, pp. 151–175, May 2010.

[22] H. Daumé III, "Frustratingly easy domain adaptation," in *Conference of the Association for Computational Linguistics (ACL)*, Prague, Czech Republic, 2007.

[23] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *European Conference on Computer Vision (ECCV)*, 2010, pp. 213–226.

[24] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification," in *Conference of the Association for Computational Linguistics (ACL)*, 2007, pp. 187–205.

[25] J. Yang, R. Yan, and A. G. Hauptmann, "Adapting SVM classifiers to data with shifted distributions," in *ICDMW '07: Proceedings of the Seventh IEEE International Conference on Data Mining Workshops*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 69–76.

[26] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines (and Other Kernel-Based Learning Methods)*. Cambridge University Press, 2000.

[27] G. C. Cawley and N. L. C. Talbot, "Preventing over-fitting during model selection via bayesian regularisation of the hyper-parameters," *J. Mach. Learn. Res.*, vol. 8, pp. 841–861, May 2007.

[28] A. Luntz and V. Brailovsky, "On estimation of characters obtained in statistical procedure of recognition (in russian)," *Techicheskaya Kibernetica*, vol. 3, 1969.

[29] T. Tommasi, F. Orabona, and B. Caputo, "Safety in numbers: Learning categories from few examples with multi model knowledge transfer," in *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2010.

[30] K. Crammer and Y. Singer, "On the algorithmic implementation of multiclass kernel-based vector machines," *J. Mach. Learn. Res.*, vol. 2, pp. 265–292, March 2002.

[31] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," http://cvxr.com/cvx, Apr. 2011.

[32] F. P. Kendall, E. K. McCreary, P. G. Provance, M. M. Rodgers, and W. Romani, *Muscles: Testing and Function, with Posture and Pain*. 530 Walnut St.Philadelphia, PA 19106-3621: Lippincott Williams & Wilkins, 2005.

[33] "Futek LMD500 medical load cell (hand)." [Online]. Available: http://www.futek.com/product.aspx?stock=FSH00125

[34] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference (Statistics: a Series of Textbooks and Monogrphs)*. Chapman and Hall/CRC, 2003.

Tatiana Tommasi received the M.Sc. degree in physics and the Dipl. degree in medical physics specialization from the University of Rome, La Sapienza, Italy, in 2004 and 2008, respectively. She is currently a Ph.D. student in electrical engineering at the École Polytechnique Fédérale de Lausanne, and Research Assistant at the Idiap Research Institute, Martigny, Switzerland. Her research interests include machine learning and computer vision with a particular focus on knowledge transfer and object categorization using multimodal information.

Francesco Orabona is a Research Assistant Professor at the Toyota Technological Institute at Chicago. His research interests are in the area of theoretically motivated and efficient learning algorithms, with emphasis on online learning, kernel methods, and computer vision. He received the MS degree in Electronic Engineering at the University of Naples Federico II, in 2003, and the PhD degree in Bioengineering and Bioelectronics at the University of Genoa, in 2007. He is (co)author of more than 30 peer-reviewed papers.

Claudio Castellini, Ph.D. received a Laurea in Biomedical Engineerings in 1998 from the University of Genova, Italy and a doctorate in Artificial Intelligence in 2005 from the University of Edinburgh, Scotland. He then spent 4.5 years as a postdoctoral researcher in the Advanced Robotics Laboratory of the University of Genova, Italy, working on machine learning applied to human sensorimotor data. Since 2009 he is a researcher at the DLR (German Aerospace Center) in the Bionics group. He is (co)author of more than 30 peer-reviewed papers.

Barbara Caputo is a senior research scientist at the Idiap Research Institute since 2006, where she leads the Cognitive visual Systems group. She received her PhD in Computer Science from the Royal Institute of Technology (KTH) in Stockholm, Sweden, in 2004. Her main research interests are in computer vision, machine learning and robotics, where Dr. Caputo has been active since 1999. As a result of her activities, Dr. Caputo has edited 4 books, and she is (co)author of more than 70 peer-reviewed papers.