# Ultrasound Imaging for Hand Prosthesis Control: a Comparative Study of Features and Classification Methods

Valerio Ortenzi[1,2], Sergio Tarantino[2], Claudio Castellini[3] and Christian Cipriani[2]

*Abstract*— **Controlling a robotic rehabilitation artefact such as a hand prosthesis is yet a rather open problem. Particularly, the choice of a human-machine interface (HMI) to enable natural control is still debatable. The traditional choice, i.e. surface electromyography (sEMG), suffers from a number of problems (electrode displacement, sweat, fatigue) which cannot be easily solved. One of its main drawbacks is the inherent low spatial resolution, at least in the standard settings. To overcome this hindrance, several novel HMIs have been proposed to substitute or augment sEMG; among them, pressure and tactile sensing, and ultrasound imaging (US).**

**In this paper we propose an advancement towards the usage of US as a HMI for hand prosthetics; namely, we compare traditional US image features with Histograms of Oriented Gradients used as input for three classifiers, and show that a high number of hand configurations and grasping force levels can be classified way above chance level by choosing the right combination of features and classifier. In an experiment involving three intact human subjects, a classification accuracy of 80% was obtained; when classifying three different levels of grip force for four grasps, the performance reduces to 60%.**

**These results confirm the usability of US imaging as a HMI for hand prosthetics, and pave the way to its practical usage as a means of natural prosthetic control.**

## I. Introduction

In the quest for ever-better Human-Machine Interfaces (HMIs) to control assistive devices, the community is trying to go beyond traditional surface electromyography (sEMG) to provide a radically better, reliable and natural form of control [1]. Novel techniques to obtain more reliable control signals from the HMIs as well as novel HMIs to substitute or augment sEMG [2] are being sought after.

This shift in paradigm motivates recent investigations of a number of such novel HMIs, including pressure sensing [3], computer vision [4] and ultrasound imaging. This latter technique in particular enforces high spatial and temporal resolution and is widely available in hospitals [5]. In contrast to sEMG, there is no reason to believe that US is influenced by sweat since the application of a gel on the skin of the patience is a common practice. Neither it is affected by muscular fatigue, since it does not provide information on force, but on position, i.e. images. Techniques exploiting US imaging are however very sensitive to the motion of the transducer with respect to the skin of the patience.

Its usage as a HMI in the context of prosthetic wrist control [6], single-finger-force detection [7], [8] and rehabilitation in a virtual-reality environment [9] has been explored with a remarkable degree of success. The existence of a simple linear relationship between finger forces and positions and gradient values referring to Regions Of Interest (ROI-gradient features) of the US images [10] encourages the exploration of new, even better ways of exploiting the wealth of information that US images carry.

In this paper we compare the image features traditionally used in this context, namely ROI-gradient features [11], with Histograms of Oriented Gradients (HOGs, [12]) in a classification scenario. While ROI-gradient features are simple averages of the grey values in the image, evaluated on a uniform grid of regions of interest, HOGs represent an image through the distribution of gradients extracted from cells and blocks constituting the image itself. By dividing the image into smaller regions, the correspondent HOGs features are local histograms of gradient directions over the pixels of these regions. All these histograms form the representation of the image.

Three intact subjects were engaged in an experiment consisting of recording US images of the transverse section of their forearm while performing ten hand postures and grasps: six postures (relaxed hand posture, and gestures as for counting from one to five), and four functional grasps (pinch, cylindrical, lateral and tripod grasp) with three levels of force.

Both sets of features, ROI-gradient and HOGs, were extracted from the US images and fed to three different classifiers, namely a Linear Discriminant Classifier (LDA) [13], [14], a Naive Bayes classifier [15] and a classifier based upon Decision Trees [15]. In the optimal setting, namely LDA on HOG features, we achieved an accuracy of 80% when classifying all the six postures and the four functional grasps with medium force level, and 60% when classifying the functional grasps with three levels of force.

These results let us hope that advanced machine learning techniques applied to US imaging will in the mid-term enter the therapists' room, to control rehabilitation devices and/or virtual-reality environments, to an outstanding degree of precision. Interestingly, and with a deeper investigation of the use of US imaging, such a HMI could potentially be used as a treatment to alleviate phantom-limb pain, which is deemed

[1] Department of Mechanical Engineering, University of Birmingham, Edgbaston, Birmingham, West Midlands B15 2TT, UK. email: vxo344@bham.ac.uk

[2] The BioRobotics Institute, Scuola Superiore Sant'Anna, Pontedera, 56025 Italy.

[3] Institute of Robotics and Mechatronics, German Aerospace Center, Oberpfaffenhofen, Germany.

978-1-4799-1808-9/15/$31.00 ©2015 IEEE

to be of cortical origin [16] and subject to change upon restoring the patient's visual feedback loop as it happens in mirror therapy [17], [18].

## II. Materials and Methods

### A. Ultrasound images acquisition

US images were collected using a portable ultrasound scanner (MyLabFive, Esaote) equipped with a LA435 linear transducer/probe and set to the B-mode (2D imaging), in order to represent the transverse section of the forearm beneath the transducer as a grey-scaled image. A custom-built plastic cradle was used to firmly hold the US transducer on the forearm, by means of an elastic band, in order to minimise motion artefacts during the tests.

Due to the position of the probe, the settings of the US scanner were adjusted in order to resemble the ones found in Castellini et al. [11], to correctly visualise the extrinsic forearm muscle: ultrasound frequency at 18MHz, focus point at a depth of 3 cm, gain at 76%, minimum depth of field ("focus number" set to 1). This resulted in a frame rate of 29 Hz. A commercial PCI frame grabber (PEXHDCAP, StartTech) was used to acquire the stream of images coming from the US scanner directly by the PC. The images were streamed through the VGA port at the standard resolution of 1024x768 pixels. The stream of images was stored on a hard disk for offline analysis.

### B. Experimental protocol

Three able bodied subjects were involved in this study. Each participant sat on a chair in front of a table with his/her forearm comfortably lying on a cushion on the table in natural position, neither pronated nor supinated (Fig. 1). Physical constraints were further placed at either sides of the forearm in order to prevent unwanted pronation/supination. The probe was thus placed on the midpoint between the wrist and the elbow on the anterior compartment of the forearm directly over the radius, with an angle of 60 deg above the horizontal. The placement of the probe was verified by visual inspection of the US images, in order to ensure that the extrinsic flexor muscles of the fingers were visible.

Each subject was verbally instructed to enact and hold 10 hand postures for two seconds. The recording of the US images started when the subject was in the selected posture/grasp and the images were stored in manually labelled folders. The postures included relaxed hand; counting from one to five; and functional grasps pinch, cylindrical, lateral and tripod grasp, each one with three levels of force (Low, Medium, High) (see Fig. 2). In the end, 18 postures were enforced. The different force levels for the functional grasps were achieved by grasping an analogue pressure gauge (North Coast Medical, Morgan Hill, CA) with specific forces (2, 3 and 5 Pound per Square Inch (psi)) and by observing the measured grip force. The 18 classes were performed in a sequence and the same sequence was repeated for 10 times, for a total of 180 trials, i.e., 2 sec x 29 Hz x 18 classes x 10 trials = 10440 images for each subject. A short pause was included between sequences in order to avoid fatigue and/or discomfort.

Fig. 1. Experimental setup used to record the US images from the three able bodied subjects.

### C. Image Processing and Classification

Two main objectives were addressed: i) to assess whether 10 different hand postures/grasps could be classified (Objective A), and ii) if grasps with different grip forces could be classified (Objective B). In turn, we split the recorded US images associated to the different classes into five datasets. The first dataset was aimed to assess the ability of recognising different hand postures/grasps (Objective A), hence it included the US images related to the six hand postures and the four functional grasps with only one level of grip force (i.e., medium) for each grasp. The second to fifth datasets were aimed to assess the ability of recognising different force levels, for each grip type (Objective B).

Three state-of-art classifiers were used to classify images from the five datasets: a linear discriminant analysis classifier (LDA) [13], [14], a Naive Bayes classifier [15] and a Decision Trees classifier [15]. The three classifiers were fed with visual features extracted from each image, in particular, ROI-gradient and HOG features. ROI-gradient features have already been proposed as a means to train linear regressors in [11] and [7]; they are triplets of $\alpha_i, \beta_i, \gamma_i$ gradient values, each triplet referring to a Region Of Interest (ROI). The ROIs are organised in a uniform grid. Intuitively, $\alpha_i$ and $\beta_i$ are the mean gradient along the $x$ and $y$ axes respectively,

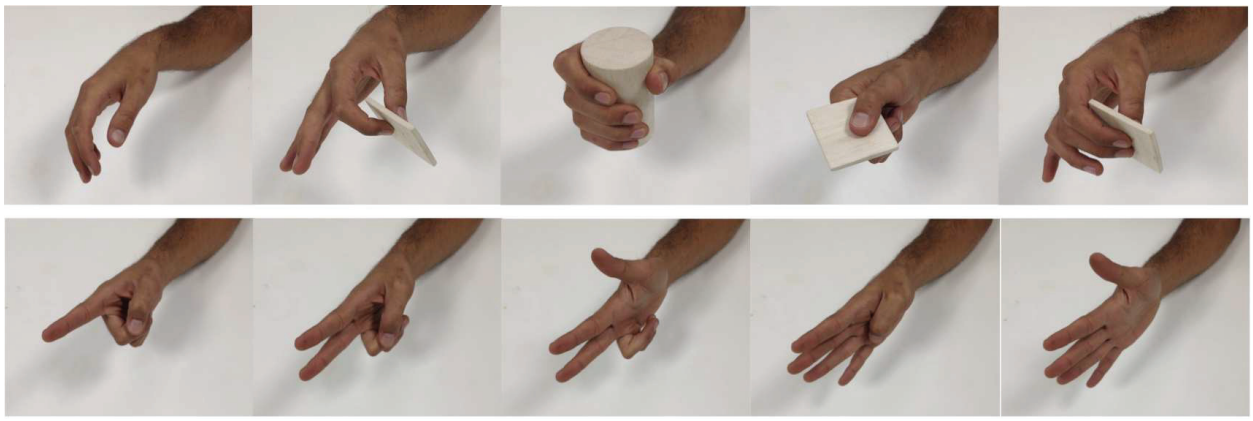

Fig. 2. Hand postures and grasps. Top row: relaxed hand posture, pinch, cylindrical, lateral and tripod grasps; bottom row: counting from one to five.

and $\gamma_i$ is an offset [11]. HOGs are also local visual features and they have been presented by Dalal and Triggs in 2005 [12]. The basic idea of HOGs is to describe an object through the distribution of gradients of grey values, extracted from cells and blocks, i.e., portions, in which the image can be fractionated into. Practically, the image is divided into smaller spatial parts, the cells, and for each cell an histogram of gradient directions is computed over the pixels belonging to the cell. The HOG features of an image are the combination of the histograms of its cells.

We extracted both sets of features from all the images of the datasets offline. We trained the classifiers on features belonging to subsets of images of the datasets. Namely, for Objective A, we ran 10 tests on an increasing number of randomly-chosen trials for each hand posture/grasp to train the classifiers from, namely from 3 to 5 out of the 10 present in each dataset. Thus, the classification problem is with ten posture/grasp classes, i.e., six hand postures plus the four functional grasps with only one level of grip force (medium), on the trials not used for training. Furthermore, we ran tests using an increasing percentage of images from every trial, from 30% to 50%, to tackle the same classification problem. In order to test the classification on different grip forces (Objective B), we used the same methodology as in the first part of Objective A, but using solely images referring to the functional grasps (pinch, cylindrical, lateral and tripod grasps) with the three levels of force (Low, Medium, High). Tests were run disjunctively on each dataset, thus disjunctively on each subject.

Matlab[1] scripts were used to extract visual features and classify them. The HOG features were extracted as proposed by Ludwig et al. in [19].

## III. Results and Discussion

Taking into account the preparation of the experimental setup, the total duration of the experiment was roughly 1 hour and 20 minutes; neither fatigue nor discomfort were reported by any of the subjects.

[1]MATLAB, version 8.3.0.532 (R2014a): The MathWorks Inc., 2014.

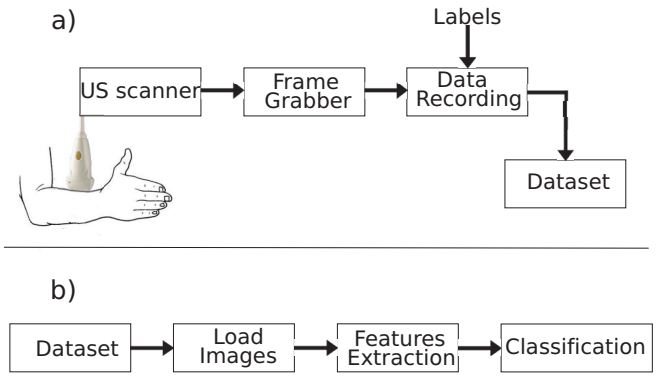


Fig. 3. Image acquisition blocks diagram. The 2D images of the forearm section under the US probe were labeled and stored in a dataset (a). In the offline analysis, the features were extracted from the images and then used for classification (b).

### A. Objective A

Classifiers using HOG features consistently outperformed those using ROI-gradient features when selections of trials were used for training (Fig. 4 and 5). The LDA classifier always outperformed the other classifiers. In particular the LDA based on HOG features outperformed the others and obtained an average performance larger than 80% (subject 1 with 3 trials in the training set). However, the Naive Bayes classifier achieved a minimum of 70% of success. The performance of Decision Trees classifiers were consistently worse than the performance of the previous two classifiers. Fig. 5 reports the average results over the three subjects in this study.

The better performance achieved with HOG features is probably due to the fact that HOG features were more numerous than ROI-gradient ones, so they better described locally each image. However, we do not think that such a difference in number might put in danger our goal to have such a classifier running online, thus classifying new images in realtime. As a matter of fact, there is almost no difference in the computational time needed for the extraction of the two sets of features. It is also true that we have not focused

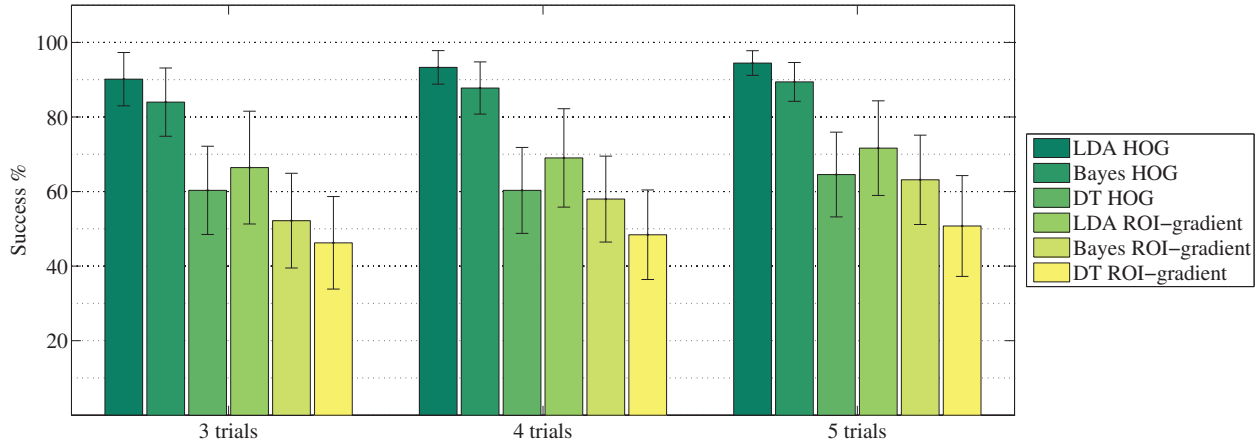


Fig. 5. Objective A: Classification performance vs. number of trials (repetitions) used for training. Results of the LDA, Bayes and Decision Trees classifiers, on subjects 1, 2 and 3. Results are averaged over the 10 classes and over the 3 subjects. Error bars denote the standard deviation.

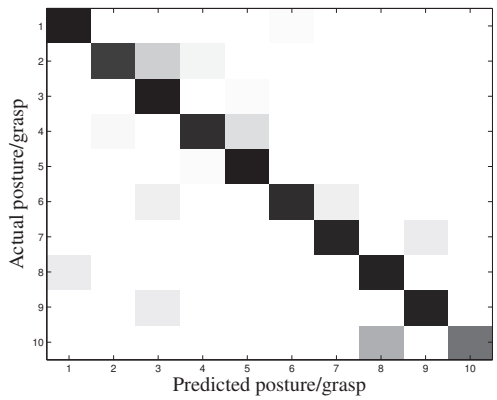


Fig. 4. Objective A: Representative confusion matrix referring to a test with the LDA classifier trained on 3 trials. The classes reported are, in turn, pinch grasp, counting-2-to-5, lateral grasp, counting-1, cylindrical grasp, rest, and tripod grasp.

on minimising the computational time so we cannot provide insights on such matter. Surely future work will revolve heavily around this attribute of the system and have stricter requirements.

When using a percentage of images from every trial, the performance of each classifier (Fig. 7) was always higher than 80%, achieving close to 100% average accuracy when using LDA on HOG features. These results are attributed to the fact that inside each trial, i.e., in each movement repetition, the images were very similar, and the reason is to be found in the method we used to obtain the images. In other words, subjects were asked to hold the hand postures and grasps for two seconds, thus the images relative to the each trial do not contain a large variability.

### B. Objective B

Similarly to Objective A, the LDA generally achieved better results with respect to the other classifiers, settling around 60%. In this case the difference with respect to the other classifiers was less substantial though. The imbalance in performance between LDA and Naive Bayes classifiers is modest and around a few percentage points, while the Decision Tree classifier performed consistently at least 5-10% worse than the previous two classifiers.

By and large, results showed that the classifiers were not well suited for classifying grasps with different level of grip forces, as demonstrated by the scarce average performance of 60% (Fig. 8 and 9). Performance was severely worse than the one achieved for Objective A. This can be attributed to the fact that the functional grasps were achieved while grasping the pressure gauge, i.e., in isometric muscle contractions. Specifically, the muscles involved in a particular grasp do not differ by changing the force grip level. On the contrary, muscles involved in different postures and grasps differ significantly, like for Objective A, and they do result in severe modifications of the transverse section of the forearm beneath the transducer. As a consequence, the resultant images have larger variance and lead to a better classification (Fig. 6). We would like to stress that there is no need for the classifiers to have any a priori knowledge about anatomy features, and this is a very important feature for this system to be used in a more general setting.

## IV. Conclusions

In this work we have proposed a comparison of two sets of ultrasound image features and three classifiers, applied to a psychophysical experiment in which three intact subjects would perform six postures (relaxed hand posture, and gestures as for counting from one to five), and four functional grasps (pinch, cylindrical, lateral and tripod grasp) with three levels of force (Low, Medium, High).

Our experimental results show that the LDA classifier trained with HOGs outperformed the others and achieved 80% of success when classifying ten postures/grasps and 60% when classifying the functional grasps with different

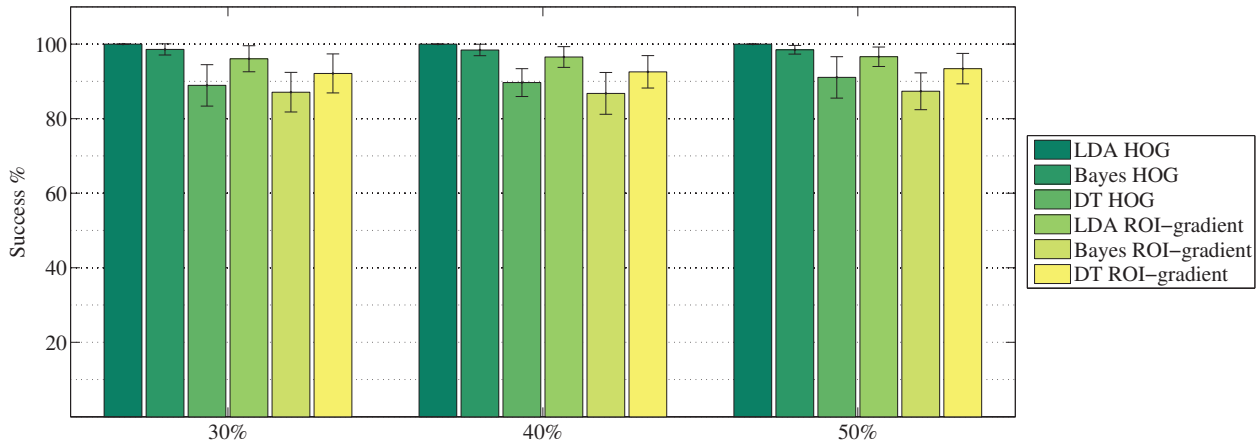


Fig. 7. Objective A: classification performance vs. percentage of images from every trial (repetition) used for training. Results of the LDA, Bayes and Decision Trees classifiers, on subjects 1, 2 and 3. Results are averaged over the 10 classes and over the 3 subjects. Error bars denote the standard deviation.

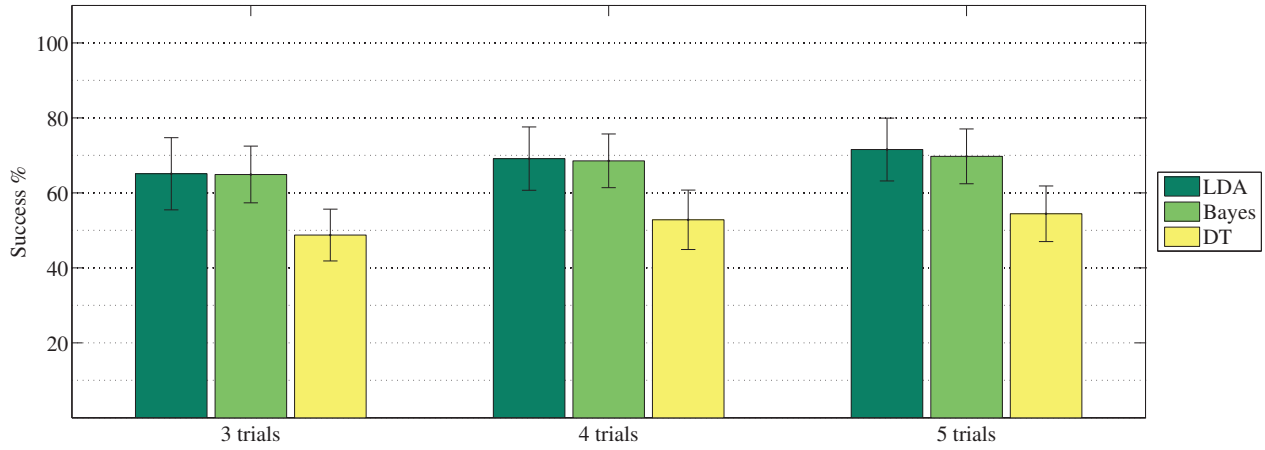


Fig. 8. Objective B): Pinch grasp, classification performance vs. number of trials (repetitions) used for training. Results of the LDA, Bayes and Decision Trees classifiers, on subjects 1, 2 and 3. Results are averaged over the 3 classes and over the 3 subjects. Error bars denote the standard deviation.

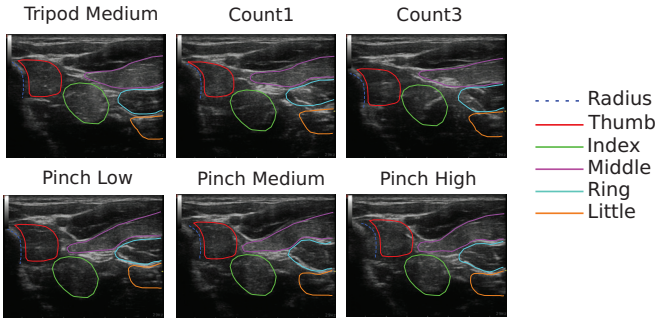


Fig. 6. Examples of US images. Top row: three hand grasp/postures, i.e., tripod grasp with medium grip force, and the gestures of counting one and three. Bottom row: same hand grasp, i.e., pinch grasp, at the three different level of grip force. The radius bone and muscles/tendons referring to the fingers were marked manually by visual inspection.

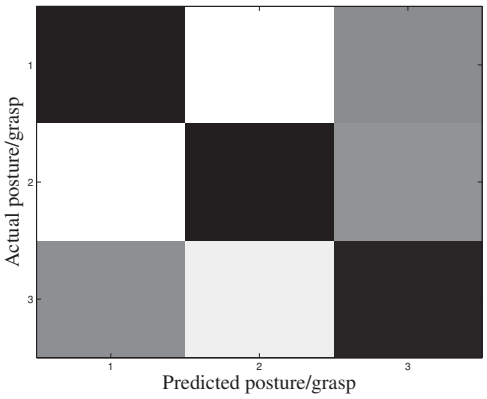


Fig. 9. Objective B: pinch grasp. Representative confusion matrix referring to a test with LDA classifier trained on 3 trials. The classes reported are in turn: high, low and medium grip force pinch grasps.

level of grip force.

These classification techniques work satisfyingly while classifying different postures/grasps, but they perform poorly on classes of different grasping forces. This is probably due

to the inherent similarity among the signals produced while enacting one single grasp but at different levels of forces. This opens up the possibility of coupling the optimal classifier with a regression method [7]: the classifier would decide which posture to activate, while the regressor would detect the intended level of force. This way a parallel position/force control system would be obtained, along the lines of [20]. All in all, the apparently unsatisfactory results shown hereby are justified by the high number of classes considered; one way of alleviating this problem is that of choosing on-the-fly the appropriate *subset* of postures required whenever the subject is facing a specific task. For instance, while withdrawing money from an ATM, the patient typically only needs a flat grip to pick the credit card, then a point-index posture to type the PIN in, then the flat grasp again to withdraw the money and the card, in this sequence.

We advocate that this technique represents a good start and we can build from here to devise a more performant solution. Besides trying more combinations between features and classifiers (e.g., Support Vector Machines [21] and Random Fourier Features [22]), future work includes devising a classification scheme that works realtime, so an analysis of the time issues will be of substantial importance; incremental learning is also a very desirable aspect, in order to increase the stability of the prediction. Research along this path has as its main application the rehabilitation of stroke patients, amputees and/or other sufferers of neuromuscular degeneration diseases. We hope that in the near future the accuracy and naturalness of control obtained using these techniques will be used in any scenario in which the intent of a patient needs to be detected and a mechanical artefact / virtual environment must be correspondingly actuated. For the very same reasons, such human machine interface may also prove useful to alleviate phantom-limb pain.

## V. Acknowledgements

This work was supported by the Italian National Workers Compensation Institute (INAIL) under the PPR3 Project.

## References

[1] N. Jiang, S. Dosen, K.-R. Müller, and D. Farina, "Myoelectric control of artificial limbs - is there a need to change focus?" *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 148–152, 2012.

[2] C. Castellini, P. Artemiadis, M. Wininger, A. Ajoudani, M. Alimusaj, A. Bicchi, B. Caputo, W. Craelius, S. Dosen, K. Englehart, D. Farina, A. Gijsberts, S. Godfrey, L. Hargrove, M. Ison, T. Kuiken, M. Markovic, P. Pilarski, R. Rupp, and E. Scheme, "Proceedings of the first workshop on peripheral machine interfaces: Going beyond traditional surface electromyography," *Frontiers in Neurorobotics*, vol. 8, no. 22, 2014.

[3] M. Wininger, N.-H. Kim, and W. Craelius, "Pressure signature of forearm as predictor of grip force," *Journal of Rehabilitation Research and Development*, vol. 45, no. 6, pp. 883–892, 2008.

[4] M. Marković, S. Došen, C. Cipriani, D. Popović, and D. Farina, "Stereovision and augmented reality for closed-loop control of grasping in hand prostheses," *Journal of Neural Engineering*, vol. 11, no. 4, p. 046001, 2014. [Online]. Available: http://stacks.iop.org/1741-2552/11/i=4/a=046001

[5] C. Castellini, "State of the art and perspectives of ultrasound imaging as a human-machine interface," in *Neuro-robotics: from brain-machine interfaces to rehabilitation robotics*, ser. Trends in augmentation of human performance, P. Artemiadis, Ed. Springer Netherlands, 2014, vol. 2, pp. 37–58.

[6] G. Jing-Yi, Z. Yong-Ping, L. P. J. Kenney, A. Bowen, D. Howard, and J. J. Canderle, "A comparative evaluation of sonomyography, electromyography, force and wrist angle in a discrete tracking task," *Ultrasound in Med. & Biol*, vol. 37, no. 6, pp. 884—891, 2011.

[7] D. Sierra González and C. Castellini, "A realistic implementation of ultrasound imaging as a human-machine interface for upper-limb amputees," *Frontiers in Neurorobotics*, vol. 7, no. 17, 2013.

[8] S. Sikdar, H. Rangwala, E. B. Eastlake, I. A. Hunt, A. J. Nelson, J. Devanathan, A. Shin, and J. J. Pancrazio, "Novel method for predicting dexterous individual finger movements by imaging muscle activity using a wearable ultrasonic system," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 22, pp. 69–76, 2014.

[9] C. Castellini, K. Hertkorn, M. Sagardia, D. Sierra González, and M. Nowak, "A virtual piano-playing environment for rehabilitation based upon ultrasound imaging," in *Proceedings of BioRob - IEEE International Conference on Biomedical Robotics and Biomechatronics*, 2014, pp. 548–554.

[10] C. Castellini and G. Passig, "Ultrasound image features of the wrist are linearly related to finger positions," in *Proceedings of IROS - International Conference on Intelligent Robots and Systems*, 2011, pp. 2108–2114.

[11] C. Castellini, G. Passig, and E. Zarka, "Using ultrasound images of the forearm to predict finger positions," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 6, pp. 788–797, 2012.

[12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *International Conference on Computer Vision & Pattern Recognition*, vol. 2, June 2005, pp. 886–893.

[13] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 7, pp. 179–188, 1936.

[14] G. J. Mclachlan, *Discriminant Analysis and Statistical Pattern Recognition (Wiley Series in Probability and Statistics)*. Wiley-Interscience, Aug. 2004.

[15] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Prentice Hall, 2009.

[16] M. Diers, C. Christmann, C. Koeppe, M. Ruf, and H. Flor, "Mirrored, imagined and executed movements differentially activate sensorimotor cortex in amputees with and without phantom limb pain," *PAIN*, vol. 149, no. 2, pp. 296 – 304, 2010.

[17] V. S. Ramachandran, D. Rogers-Ramachandran, and S. Cobb, "Touching the phantom limb," *Nature*, vol. 377, no. 6549, pp. 489–490, 1995.

[18] B. L. Chan, R. Witt, A. P. Charrow, A. Magee, R. Howard, P. F. Pasquina, K. M. Heilman, and J. W. Tsao, "Mirror therapy for phantom limb pain," *New England Journal of Medicine*, vol. 357, no. 21, pp. 2206–2207, 2007. [Online]. Available: http://www.nejm.org/doi/full/10.1056/NEJMc071927

[19] O. Ludwig, D. Delgado, V. Goncalves, and U. Nunes, "Trainable classifier-fusion schemes: An application to pedestrian detection," in *12th International IEEE Conference on Intelligent Transportation Systems, 2009. ITSC '09*, 2009.

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

[21] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory (COLT)*, D. Haussler, Ed. ACM press, 1992, pp. 144–152.

[22] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Advances in Neural Information Processing Systems 20*, 2008, pp. 1177–1184.