Comparing Recognition Methods to Identify Different Types of Grasps for Hand Rehabilitation

Beatriz Leon, Angelo Basteris and Farshid Amirabdollahian
Adaptive Systems Research Group
University of Hertfordshire
Hatfield, Hertfordshire, UK
Email: b.leon, a.basters, f.amirabdollahian2@herts.ac.uk

***Abstract*—Grasping activities are extremely frequent in the set of activities of daily living. This causes severe impairments for stroke survivors, whose wrist and hand may suffer from a variety of symptoms such as spasticity, hypertone and muscular weakness. Intensive repeated movement performance is at the base of robot-therapy. Thus, patients may benefit, in terms of functional recovery, from the integration of grasp gestures in robot mediated exergaming. In this feasibility study, we developed and tested three methods for recognizing four different grasp postures performed while wearing an exoskeleton for hand and wrist rehabilitation after stroke. The three methods were based on the statistics of the produced postures, on neural networks and on support vector machines. The experiment was conducted with healthy subjects, with no previous injuries on the hand, during grasping of actual objects and then repeated using imaginary objects. We compared the three methods in terms of accuracy, robustness with respect to the size of the training sample, inter-subjects' variability, differences between different postures and evaluating the presence of real objects. Our results show that the support vector machine method is preferable in terms of both accuracy and robustness, even with a small training sample, with training times on the order of milliseconds.**

Keywords*—grasp posture recognition; stroke rehabilitation; Support Vector Machines; Neural Networks.

I. Introduction

Stroke survivors are often unable to perform tasks requiring fine motor control, which are needed during activities of daily living, among which grasping is one of the most recurrent. Robots represent an excellent tool for exercise-based approach in neurorehabilitation, but in order to increase the functional outcome of the treatment patients training should consist in the performance of grasping movements similar to those needed in daily life [1].

The Supervised Care & Rehabilitation Involving Personal Tele-robotics (SCRIPT) project aims at delivering an affordable system for home-based rehabilitation of the hand and wrist for stroke survivors [2]. An exoskeleton has been developed within the project in order to facilitate the patients' fingers and wrist extension, due to observed abnormal flexion synergies often leading to flexed hand and wrist [3]. The motivation of the participant is enhanced by therapeutic exercises mediated via interactive games, which subjects control by wearing the orthosis and performing several movements of the whole upper limb, from the shoulder to the fingers.

There are numerous types of grasping. Feix et al. refer to 17 gross types of grasping [4]. These are utilised within the daily life interaction. Our aim is to incorporate the detection of these grasps in the rehabilitation framework so that they can be incorporated into human-machine interaction with the aim of increasing interaction time.

The problem of hand posture detection has been approached with several methods. A preliminary gross distinction can be made between vision-based or glove-based approaches. Vision based approaches allow natural hand gesture recognition, typically by detection of the fingertips. Detailed information about such type of technologies can be found on comprehensive reviews such as Chaudary et al. [5]. Among the vision-based techniques, a possibility is to use specific color patterns on a fabric glove in order to facilitate the detection [6]. Glove-based approaches reduce the need for computational power at the cost of possibly altering the natural performance of the movements by making one wear an instrumented glove. This could not be a concern when the presence of such a device is however required in order to assist the patient in movement performance. Several methods have been proposed for hand posture detection using data gloves [5], [7], including feature extraction (e.g., principal component analysis, linear discriminant analysis) and learning algorithms (e.g., neural networks, support vector machines). Specifically, Xu et al. [8] allowed hand gesture recognition for driving by training and testing a neural network with a pattern of 300 hand gestures. Other studies have also compared different approaches for grasp detection. Palm et al. [9] compared three methods: difference between the grasp and models grasps, qualitative fuzzy models and Hidden Markov Models (HMM), concluding that the first method outperforms the other two in terms of accuracy on a set of 10 grasps for 15 different grasps primitives. All the aforementioned studies used a data glove, which measures the angles of 18 joints in the hand (CyberGlove [10]).

In the context of this study, the vision-based approaches are unsuitable given that the exoskeleton causes the visual occlusion of most of the hand. Therefore, the gesture recognition should rely on the sensor readings provided by the device. Additionally, a basic requirement for facilitating the application of grasp recognition in the rehabilitation framework, possibly affecting the future use of the system, is a short setup time.

Another aspect relates to the involvement of real or imag-

 ISBN: 978-1-61208-325-4

inary objects while performing the grasp postures. Ideally, patients should mimic the graping posture without interaction with real objects while playing the games. However, there could be substantial differences between the hand posture observed when holding actual objects and its performance based on motor imagery, which can affect the therapy outcome. On the one hand, the use of real objects would possibly enhance the skills transfer from the training activity to the functional use. Also, grasping real objects would induce physical constraints on the hand posture, which might facilitate the gesture recognition, making its performance more repeatable. However, having real objects also carries disadvantages such as introducing additional requirements in the device design, having the participants focusing their gaze and attention on the objects rather than on the screen and reducing the usability of the system.

Hence, in this work, we test the trade-off between accuracy and training sample size of three different approaches: recognition based on statistics, based on neural networks and based on support vector machines, for three type of grasps, either using actual objects or imagined objects while performing the gestures.

The remainder of the paper is structured as follows. We start with the methods section introducing the passive orthosis used in the SCRIPT project, the selected grasp postures, the different methods selected to recognize the postures and details of the experimental protocol. Then, we report and analyze the results, and conclude with a brief summary of our findings, including the best method found for grasp posture recognition and directions for future work.

II. Methods

A. Measuring device

The SCRIPT passive orthosis [11] is a passive device which features five leaf springs, which are connected to finger caps through an elastic cord (Figure 1). The extension forces resulting from the parallel of these two elements are applied at the fingertip. The elastic cord enables the fingers freedom of movement relative to the leaf spring and also allows to adjust the level of support provided by the device. The leaf springs are fitted with commercial resistive flex sensors [12], which measure their deflexion with a resolution of 1 degree. Because of the elastic coupling, the deflection of the leaf spring is not the actual flexion angle of the finger. However, the two quantities are related by a monotonically increasing function [11]. Also, movements of lateral abduction/adduction of the fingers and opposition of the thumb are not restricted nor sensed by the orthosis. It measures only overall finger flexion in a range from 0 to 90 degrees and wrist flexion and extension in a range from 90 to -45 degrees.

B. Hand Gestures

We selected three types of grasps shown in Figure 2. Two are classified as precision grips: the three-jawed chuck (the thumb opposes the index and middle finger) and the lateral prehension (the thumb holds an object against the side of

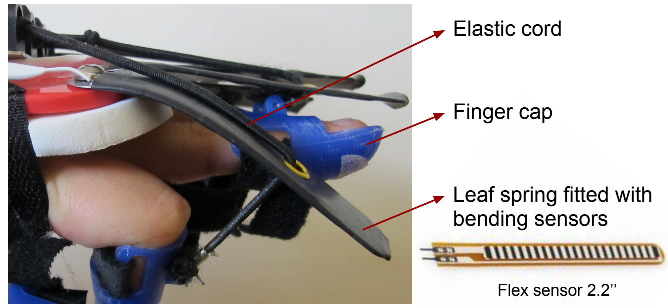

Fig. 1. Bending sensors and the leaf springs of the SCRIPT passive orthosis

the index finger). The third grasp selected is classified as a power grasp: the cylindrical prehension in which all fingers make contact with the object. Keller et al. [13] identified the three-jawed chuck and the lateral prehension as the most frequently used prehensile patterns for static and dynamic grasping, respectively. Additionally, the three-jawed chuck and the cylindrical prehension are tested as part of the Grasp and Grip sections of the Action Research Arm Test [14], which has been used as a reliable, valid measure of arm motor status after stroke. The relaxed posture of the hand was used as the forth gesture in order to be able to recognize when the patients are not performing any grasps. Furthermore, these gestures were selected considering that patients with different levels of hand impairment should be able to perform them.

C. Methods for Recognition

1) Recognition based on statistics of the training samples: The method considers the absolute error of each finger with respect to the average value of the training samples. In the training phase, the mean value of the flexion measured for each finger f is calculated for the N number of training samples:

$$mean_f = \frac{\sum_{i=1}^{N} |Flexion_f i|}{N} \tag{1}$$

During the testing phase, the flexion values of each finger are compared with the mean value of the training phase and averaging it among the five fingers, identifying then the gesture only if this value falls below a threshold th:

$$\frac{\sum_{f=1}^{5} |Flexion_f - mean_f|}{5} \leq th \tag{2}$$

The value of th was empirically set to 10 degrees as this value allowed to reach an accuracy of 90% when tested on a single subject during a pilot study.

With this method, gestures might be recognized even though one or more fingers are in very different conditions from what measured in the training phase, provided that other fingers compensate for this overall difference.

2) Recognition based on Neural Networks (NN): Artificial Neural Networks (NN) [15] have been extensively used for supervised learning to solve problems of pattern recognition (classification) and regression [16]. They have been previously used for hand posture classification [5], [17].

 ISBN: 978-1-61208-325-4

(a) Three-jawed chuck (b) Lateral prehension (c) Cylindrical prehension (d) Relaxed posture

Fig. 2. Selected gestures to be recognized. Bottom images show how they are performed wearing the SCRIPT device.

In this work, we used a three layer neural network that contain five input nodes for the flexion values of each finger, 10 hidden nodes and four output nodes, one for each gesture to be recognized (the three selected gestures plus one relaxed postured acquired by instructing the subjects to relax their fingers). The flexion angles were normalized in a range from 0 to 1, corresponding to 0 to 90 degrees. Similarly, the limits of the output nodes were set to 0 and 1. We used back propagation, a learning mechanism that minimizes the error between target output and the output produced by the network until it reached an error of 0.01.

After the training phase, a gesture was recognized if the results given by an output node was higher than 0.7 and the other gestures has a recognition rate less than 0.3. Otherwise, no gesture was returned by the model for the given posture. This method was implemented in Python using the *Neurolab* library [18].

3) Recognition based on Support Vector Machines: This method utilises Support Vector Machines (SVM), which is a popular machine learning technique for classification [19]. A support vector machine constructs a set of hyperplanes in a high-dimensional space that are used to classify the data. A good separation is archived by the hyperplane that has the largest distance to the nearest training data point of any class. The hyperplanes are found solving the following optimization problem [20]:

$$\min_{w,b,\xi} \quad \frac{1}{2}w^T w + C\sum_{i=1}^{l}\xi_i \tag{3}$$
$$\textit{subject to} \quad y_i(w^T\phi(x_i) + b \geq 1 - \xi_i) \textit{ and } \xi_i \geq 0$$

where $\{(x_i, y_i)|x_i \in R^p, y_i \in -1, 1\}$ are the training set of l instance-label pairs, x_i is p-dimensional real vector, w the normal vector of the hyperplane and $C > 0$ the penalty parameter of the error term. The training vectors x_i are mapped into a p-dimensional spaces by the function ϕ and in order to create nonlinear classifiers a kernel function is used. In our work, we used a radial basis function (RBF) as the kernel function, given that it can handle the case when the relation between class labels and attributes is non-linear. It is defined as:

$$K(x_i, x_j) = exp(-\gamma \parallel x_i - x_j \parallel^2), \gamma > 0 \tag{4}$$

where γ is a kernel parameter.

Therefore, two parameters are needed: C and γ. In order to find the best values for this parameters, we used a $v - fold$ cross-validation technique, dividing the training set for one subject into v subsets of equal size and testing the classifier on the remaining $v-1$ subsets. The values found were: $C = 4$ and $\gamma = 1$. The method was implemented in Python using the *LIBSVM* package [20]. As with the previous method, the flexion angles were normalized in the range from 0 to 1. The selected error to stop the training face was 0.01.

D. Experimental protocol

This work was designed as a feasibility study aimed at comparing different methods for grasp posture recognition and selecting one that can be further adopted in rehabilitation. We decided to perform the experiments of this study on healthy subjects, while focusing on the recognition capabilities of the different methods.

Five healthy subjects (age = 31 $\pm$ 2.1, 3 males/2 females) with no previous injuries of fingers, hand or wrist volunteered to participate in this study. Participants were recruited amongst faculty staff by advertising on an internal mailing list. All subjects were right handed.

This study was carried out at the University of Hertfordshire and approved by the university ethics committee (Ethics protocol number COM/ST/UH/00008).

The participants were asked to wear a medium sized left hand SCRIPT passive orthosis while sitting in front of a PC. Participants were instructed to grasp one out of three objects (a ball, a mug and a notepad) by showing on the screen the picture of the appropriate grasp (Figure 2). The subject then confirmed with a click that he/she achieved the desired gesture and the flexion angles of the fingers were saved. After confirmation, they were asked to release the object, relax the hand and press a button. At that moment, the flexion angles of the fingers of the relaxed posture were also saved.

Each subject performed six repetitions of each gesture in a pseudo-random sequence grasping the real objects. Subsequently, participants repeated the same procedure but mimicking the required gesture without actual objects (the difference is shown in Figure 3 for the first grasp).

Data were then post-processed by Python ad-hoc applications implementing the three methods. Each technique was evaluated based on its computational time and accuracy of recognition, defined as the total number of correctly classified gestures over the total number of gestures.

The data acquired for each subject was divided into two sets for training and testing purposes. The number of training samples (N) was varied between 2 and 5 samples per gesture

Fig. 3. Example of performing a gesture while: a) grasping a real object or b) grasping an imaginary object

type. The training set was then used to train the model according to each method, which allows us to provide insights into what is the minimum number of samples required to achieve a given accuracy. Results were considered taking into account all possible permutations of the training samples, for each training sample size N. Data analysis was done using IBM SPSS for Windows version 21.0.

III. Results

A. Recognition methods

Figure 4 shows the results in terms of accuracy for the three methods, averaging all subjects. Regardless of the training sample size, the SVM approach outperforms the other methods in terms of accuracy, with median values greater than 90% already with two repetitions of a gesture. While the method based on statistics did not allow achieving comparable accuracies, neural networks can potentially be used. However, this would happen at the cost of increasing the training sample.

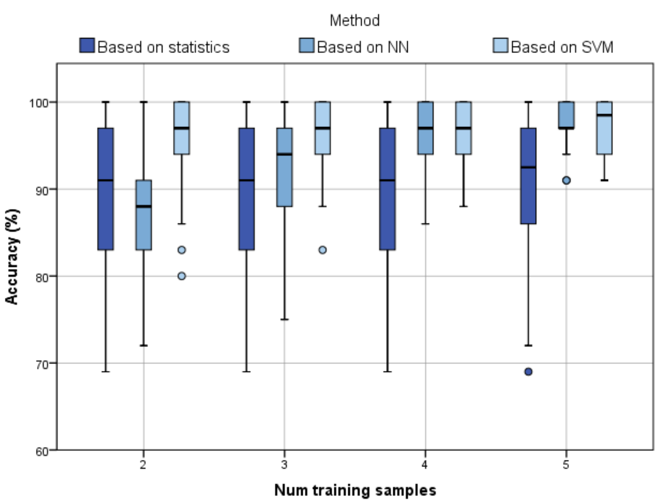

Fig. 4. Accuracy for each method tested over all subjects.

The computational times taken for each method is presented in Table I. The method based on statistics took less than 1 ms regardless of the training sample size, given that it involves only simple arithmetical and logical operations. Similarly, the method based on SVM needed a small training time (below 56 ms for all conditions). On the contrary, the method based on NN required much higher durations, ranging from 0.1 s to several hours.

TABLE I
TIME REQUIRED TO TRAIN THE DIFFERENT METHODS

	Computational time (sec)		
Method	Mean	Maximum	Minimum
Based on statistics	<0.001	<0.001	<0.001
Based on NN	89.543	>10000	0.110
Based on SVM	0.005	0.056	0.001

Therefore, although the method based on NN was able to reach median recognition accuracies higher than 90% with more than 3 training samples, the long training time, and particularly its high variability, suggested to drop it in favour of the method based on SVM, which was used to perform the further analyses evaluating the presence of real objects and the variability among subjects.

B. Grasping real vs. imaginary objects

The results of comparing the accuracy of gesture recognition using real or imaginary objects are shown in Figure 5 for the method based on SVM. The gestures performed without objects are systematically less recognized than the ones using objects.

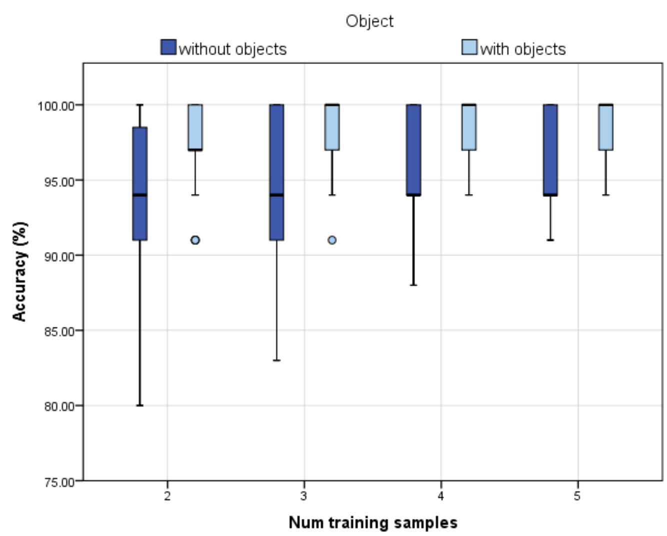

Fig. 5. Comparison of the accuracy of gesture recognition when grasping objects or imagining them, using the method based on SVMs.

This result was expected as the gestures grasping a real object are performed in a more consistent manner than when the objects are imagined. However, performing the gestures using real objects while playing a game is not practical, therefore the use of imaginary objects is advisable as the median accuracies of recognition are no lower than 90%.

C. Number of training samples required

It is important to determine the number of required training samples needed to achieve a good accuracy of recognition.

 ISBN: 978-1-61208-325-4

As the patients need to perform the training procedure before starting the games, the less number of samples required to train the model the better.

For the method based on SVMs, Figure 5 shows the results of the accuracy for different number of training samples acquired per gesture. The results show that the accuracy of recognition increases with the number of training data, as expected. We performed a contrast test between the different number of samples used for training (Table II). The results show that for grasping both real and imaginary objects the means of the groups are significantly different (<0.05) and that there is no significant difference increasing from 4 to 5 samples ($p > 0.05$), therefore we recommend using 4 samples per each gesture.

TABLE II
CONTRAST TEST OF THE EFFECT OF THE NUMBER OF TRAINING DATA OVER THE GESTURE RECOGNITION ACCURACY

Object	Means of groups significance	Contrast significance		
		2 vs. 5 samples	3 vs. 5 samples	4 vs. 5 samples
With	0.007	0.004	0.048	**0.375**
Without	<0.001	0.000	0.004	**0.174**

D. Variation of recognition per gesture

In this section, we considered the variation of accuracy among gestures. The results of the accuracy using the method based on SVMs and 4 samples for training are shown in Figure 6. It can be seen that the relaxed posture is always correctly recognized. The three-jawed chuck has also very high accuracy, despite the presence of a few outliers. The cylindrical and lateral prehension appear as those more difficult to be recognized, specially when using imaginary objects. This is likely because of their similarity, particularly enhanced by the constraints on movements imposed by the orthosis.

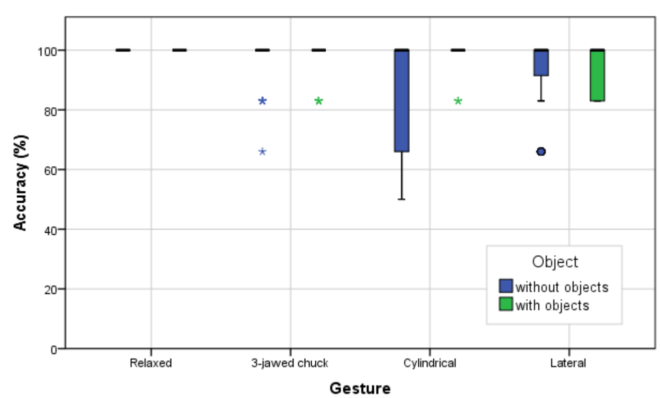

Fig. 6. Comparison of the recognition accuracy obtained for each of the selected gestures using the method based on SVMs.

E. Variation between subjects

In order to study the variation between subjects, we have first analyzed the differences in the flexion angles for each of the gestures performed by the 5 subjects while imagining the objects and then we analyzed how this variation on the flexion angles inter subjects affected the gesture recognition.

The finger flexion angles for each subject varied in most of the cases less that 12 degrees, except for the small finger that presents variation up to 21 degrees performing the lateral prehension for subject 3 and the thumb for subject 5 with a standard deviation of 19 degrees. Figure 7 shows a summary of these results over all subjects per grasp gesture. As it can be expected, the relaxed posture is the most consistent over all the gestures. For the 3-jawed chuck, the ring and small fingers present high variation (> 60 degrees) as they were not directly involved in the gesture as well as the thumb, which flexion reading can vary according to the abduction position which is not measured by the device. The cylindrical prehension shows approximately the same variation over all fingers and the lateral prehension show the highest variation since the flexion of the fingers is not highly constrained by the grasp, as long as all the fingers are flexed and the thumb is above the index finger. These results are correlated with the recognition accuracies of each gesture presented in Figure 6, which shows that the cylindrical and lateral prehension could have low recognition accuracies (up to 50%) given their high variation inter-subjects.

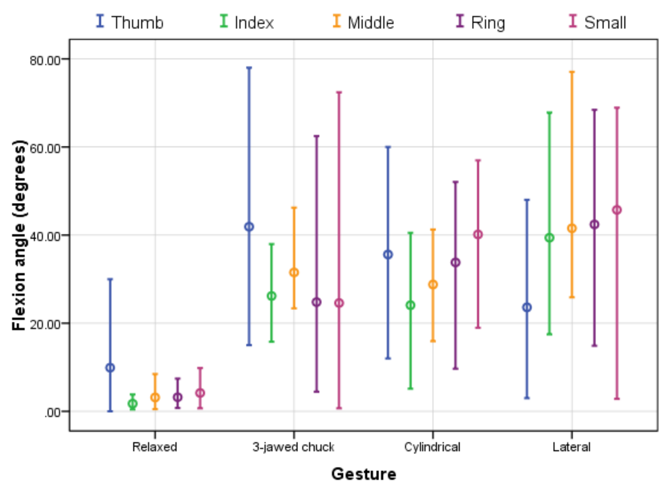

Fig. 7. Variability of the different finger flexion produced by all subjects while performing the different gestures

Finally, Figure 8 presents the accuracy of recognition for each subject (using SVM, without objects). The results shows small variation between subjects and an accuracy higher than 90% for all of them using 3 or more training data per gesture.

IV. CONCLUSION AND FUTURE WORK

In this paper, we evaluated three different methods to classify hand gestures while wearing a passive orthosis capable of measuring the angle of flexion of the fingers. By using support vector machines we could achieve an overall accuracy of more than 90%, with this methods being preferable to statistics and neural networks because of recognition accuracy and computational time.

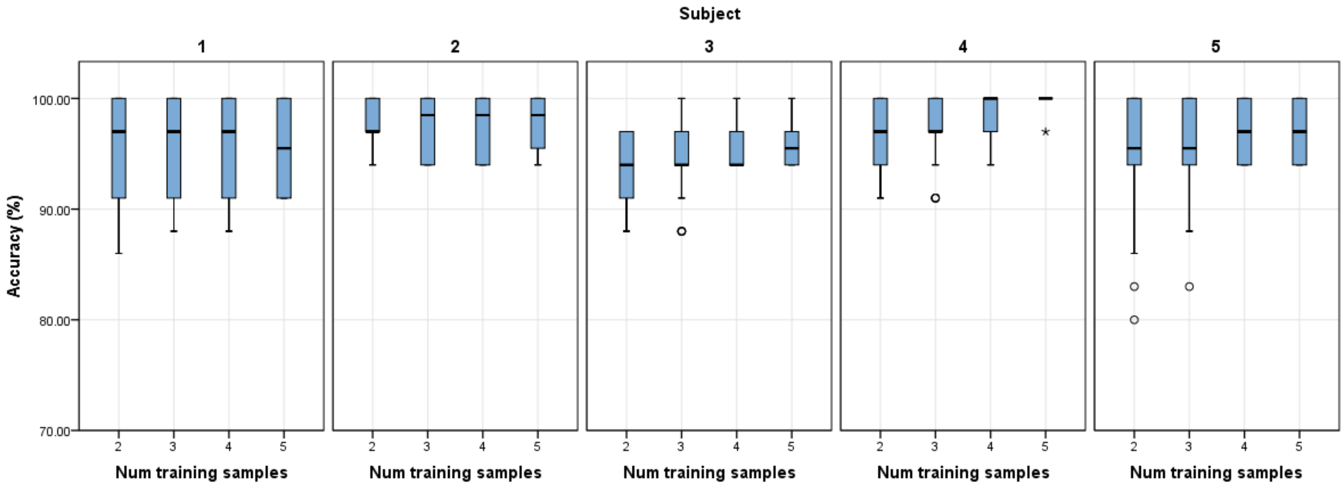

Fig. 8. Accuracy for each subject using the method based on SVMs.

This accuracy can be achieved already with only 4 training samples, thus allowing very short preparation time and making this approach suitable for home-based rehabilitation. In this sense, we also showed that the the fact that subjects pretending to grasp objects (instead of actually grasping them) had small effects on the recognition rate, suggesting that the proposed approach can work with imagined postures.

In future work, we will evaluate the performance of the selected method with data collected for post-stroke patients and compare the results with this study using healthy participants. Additionally, the set of grasps in this study was kept small in order to highlight the suitability of different algorithms, but the method could be tested including a larger number of grasps.

ACKNOWLEDGEMENTS

This work has been partly funded by the European Community Seventh Framework Programme under Grant agreement no. FP7-ICT-288698 (SCRIPT).

REFERENCES

[1] A. A. Timmermans, H. A. Seelen, R. D. Willmann, and H. Kingma, “Technology-assisted training of arm-hand skills in stroke: concepts on reacquisition of motor control and therapist guidelines for rehabilitation technology design,” J Neuroeng Rehabil, vol. 6, 2009, pp. 1–18.

[2] G. B. Prange, H. J. Hermens, J. Schafer, N. Nasr, G. Mountain, A. H. A. Stienen, and F. Amirabdollahian, “Script: Tele-robotics at home - functional architecture and clinical application,” in 6th International Symposium on E-Health Services and Technologies (EHST), Geneva, Switzerland, 2012.

[3] J. P. Dewald and R. F. Beer, “Abnormal joint torque patterns in the paretic upper limb of subjects with hemiparesis,” Muscle Nerve, vol. 24, no. 2, Feb 2001, pp. 273–283.

[4] T. Feix, H. bodo Schmiedmayer, J. Romero, and D. Kragic, “A comprehensive grasp taxonomy,” in Robotics, Science and systems conference: Workshop on understanding the Human Hand for advancing robotic manipulation, 2009, pp. 1–2.

[5] A. Chaudhary, J. L. Raheja, K. Das, and S. Raheja, “Intelligent approaches to interact with machines using hand gesture recognition in natural way: A survey,” International Journal of Computer Science and Engineering Survey (IJCSES), vol. 2, no. 1, 2011, pp. 122–133.

[6] L. Lamberti and F. Camastra, “Handy: A real-time three color glove-based gesture recognizer with learning vector quantization,” Expert Systems with Applications, vol. 39, no. 12, 2012, pp. 10 489 – 10 494.

[7] J. LaViola, “A survey of hand posture and gesture recognition techniques and technology,” Brown University, Tech. Rep. CS-99-11, April June 1999.

[8] D. Xu, “A neural network approach for hand gesture recognition in virtual reality driving training system of spg,” in Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, vol. 3, 2006, pp. 519–522.

[9] R. Palm and B. Iliev, “Grasp recognition by time-clustering, fuzzy modeling, and hidden markov models (hmm) - a comparative study,” in Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). IEEE International Conference on, 2008, pp. 599–605.

[10] C. Systems, “CyberGlove,” http://www.cyberglovesystems.com/, [retrived: January, 2014].

[11] S. Ates, J. Lobo-Prat, P. Lammertse, H. van der Kooij, and A. Stienen, “Script passive orthosis: Design and technical evaluation of the wrist and hand orthosis for rehabilitation training at home,” in ICORR13, 13th International Conference on Rehabilitation Robotics, Seattle, Washington, 2013, pp. 1–6.

[12] S. Electronics, “Flex Sensor 2.2”,” https://www.sparkfun.com/products/10264, [retrived: January, 2014].

[13] A. D. Keller, C. L. Taylor, and V. Zahm, Studies to determine the functional requirements for hand and arm prosthesis. Dept. of Engineering, University of California, 1947.

[14] N. Yozbatiran, L. Der-Yeghiaian, and S. C. Cramer, “A standardized approach to performing the action research arm test,” Neurorehabil Neural Repair, vol. 22, no. 1, 2008, pp. 78–90.

[15] J. A. Anderson, An introduction to neural networks. Cambridge, MA: MIT Press, 1995.

[16] P. Melin and O. Castillo, “Supervised learning neural networks,” in Hybrid Intelligent Systems for Pattern Recognition Using Soft Computing, ser. Studies in Fuzziness and Soft Computing. Springer Berlin Heidelberg, 2005, vol. 172, pp. 55–83.

[17] H. Hasan and S. Abdul-Kareem, “Static hand gesture recognition using neural networks,” Artificial Intelligence Review, 2012, pp. 1–35.

[18] “Neurolab Library,” https://code.google.com/p/neurolab/, [retrived: January, 2014].

[19] C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, no. 3, 1995, pp. 273–297.

[20] C.-C. Chang and C.-J. Lin, “Libsvm: A library for support vector machines,” ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, May 2011, pp. 27:1–27:27.

 ISBN: 978-1-61208-325-4