Visual Cues to Improve Myoelectric Control of Upper Limb Prostheses

Andrea Gigli\textsuperscript{1}, Arjan Gijsberts\textsuperscript{1}, Valentina Gregori\textsuperscript{1}, Matteo Cognolato\textsuperscript{2}, Manfredo Atzori\textsuperscript{2}, and Barbara Caputo\textsuperscript{1}

Abstract—The instability of myoelectric signals over time complicates their use to control highly articulated prostheses. To address this problem, studies have tried to combine surface electromyography with modalities that are less affected by the amputation and environment, such as accelerometry or gaze information. In the latter case, the hypothesis is that a subject looks at the object he or she intends to manipulate and that knowing this object’s affordances allows to constrain the set of possible grasps. In this paper, we develop an automated way to detect stable fixations and show that gaze information is indeed helpful in predicting hand movements. In our multimodal approach, we automatically detect stable gazes and segment an object of interest around the subject’s fixation in the visual frame. The patch extracted around this object is subsequently fed through an off-the-shelf deep convolutional neural network to obtain a high level feature representation, which is then combined with traditional surface electromyography in the classification stage. Tests have been performed on a dataset acquired from five intact subjects who performed ten types of grasps on various objects as well as in a functional setting. They show that the addition of gaze information increases the classification accuracy considerably. Further analysis demonstrates that this improvement is consistent for all grasps and concentrated during the movement onset and offset.

I. INTRODUCTION

The loss of a hand or an arm due to amputation has a drastic impact on the quality of life. Although advanced myoelectric prostheses have the potential to restore some of the lost functionality, their acceptance among amputees is very low \textsuperscript{1}. Aside from high cost, one of the problems with these active prostheses is that their control is not robust and requires a long and painful training procedure. Myoelectric signals change over time, for instance due to electrode shift, user adaptation, and fatigue, and this hurts control robustness.

Academic efforts have therefore started to focus on how to make prosthetic control more stable and more intuitive. An interesting avenue is to reduce the dependency on surface electromyography (sEMG) by including other sources of contextual information, such as inertial sensors \textsuperscript{2} or computer vision \textsuperscript{3, 4}. The working principle is that this context is helpful in decoding the intent of the prosthesis user. This seems obvious in case of the orientation of the limb (cf. inertial sensors), but also the user’s gaze behavior and the visual description of an object of interest may contain important side-information to determine the desired hand movement. For example, it seems more likely that a person that is fixing a pen lying on a table desires to perform a writing tripod than a power disk grasp.

Recent studies have investigated the use of visual information to preshape a prosthesis based on the estimated object size and orientation. Rather than object size and orientation, we argue that also the object’s affordances are relevant to determine the desired grasp type. We therefore extract high-level features of the object of interest using a powerful, off-the-shelf convolution neural network. These features are highly discriminative for object identification and they will therefore also contain informative content on the object’s functionality. Furthermore, in contrast to earlier studies we do not require users to trigger the visual recognition system, but instead use gaze tracking to automatically detect stable fixations and to segment the object of interest.

The proposed method was evaluated offline on data collected from five intact subjects performing ten grasps. All grasps were repeated both seated and standing, and with three different objects each. The chosen objects are associated with activities of daily living and thus representative of a home environment. To promote variability in the arm dynamics and visual scene, subjects also performed these grasps as part of 15 functional movements (e.g., open a zipper using a lateral grasp).

The remainder of this paper continues with an overview of related work in section II. In section III, we give a detailed description of our method to automatically detect fixations and how to integrate the object’s visual representation with sEMG. We then describe the experimental setup of our evaluation in section IV and follow this with the results in section V. This paper is concluded in section VI.

II. RELATED WORK

The difficulty of reliably measuring and interpreting sEMG has led to active research on the inclusion of other types of sensory modalities to control myoelectric prostheses, such as sonomyography, mechanomyography, and force myography (for detailed overviews, see \textsuperscript{5, 6}). Besides those that measure muscular activity, also modalities that provide an informative context on the intended movement have been combined with sEMG. Several studies have shown that accelerometry of the relevant arm provides useful information on arm orientation and dynamics that is complementary to sEMG \textsuperscript{2, 7}.

More recently, also computer vision and gaze information have been considered to improve intent recognition. Their
relevance has been shown in early studies, in which innovative systems were proposed for controlling the prehension of a transradial prosthesis. These either used a webcam \[3, 4\] or electro-oculography \[8\] to automatically preshape the prosthesis based on the estimated object size and orientation. This approach has subsequently been integrated with a myoelectric control strategy by Markovic et al. \[9, 10\]. In their system, myoelectric activity is combined with computer vision and inertial sensing to provide artificial proprioceptive feedback on the grasp type and object size. Via sEMG-based sequential and proportional control, the user can override the automated preselections of the system. The use of computer vision in the context of prosthetics was also hinted at by Ghazaei et al. \[11\], who used deep learning to classify grasps based on the object’s appearance.

Slightly different from our application in prosthetics for amputees, sEMG and gaze information has also been used to operate a robot arm for tetraplegic patients. Corbett et al. \[12\] use the subject’s gaze to help to determine the target position of a reaching movement, while McMullen et al. \[13\] combine this with computer vision to initiate and automatically perform the reach-grasp-drop motion of the robot arm.

## III. GAZE INTEGRATION

The basic idea behind this work is to extract a representation of the object that is observed during a prehension and use it as an auxiliary cue in support of a standard sEMG based grasp classifier. To do so, we designed a method to automatically detect stable gaze fixations, extract relevant visual information associated with those fixations, and subsequently integrate this information in the movement classifier.

### A. Fixation detection

The first step of the algorithm consists in finding fixations in the gaze tracking data. A fixation consists of a period of time (generally between 350 ms to 450 ms \[14\]) where the eye-gaze remains in a limited area of the visual field. Since we are only interested in fixations that precede a grasp, we attempt to identify an increase in muscle activity by looking at the Root Mean Square (RMS) of the myoelectric signals in a sliding window of length \(\tau\). As can be seen in Figure 1, the average activity over all electrodes, denoted with \(\text{AvgRmsEmg}\), increases drastically during the initial reach-to-grasp phase.

We identify these increases in an online manner using Bollinger bands, which calculate the number of standard deviations that a current value \(x\) deviates from a historical mean within a sliding window of length \(\tau\) and \(\eta\):

\[
b(x, t, \tau_{boll}) = \frac{x_t - \mu(x_{t:t-\tau_{boll}})}{\sigma(x_{t:t-\tau_{boll}})},
\]

where \(x_{t:t-\tau_{boll}}\) denotes the sliding window mean and standard deviation. Since we are interested in sudden increases in muscle activity, we limit our attention to when the value exceeds the upper Bollinger band

\[
b(x, t, \tau_{boll}) \geq \eta,
\]

where \(\eta\) regulates the sensitivity of the method to the signal’s variations. Figure 1 shows the identified time intervals associated to the grasp’s onset (red) when \(\text{AvgRmsEmg}\) exceeds its upper Bollinger band (green).

After identifying these regions where the hand has started reaching, we identify stable fixations on the basis of the gaze volatility. Since the gaze is represented as a 2-dimensional vector (both \(x\) and \(y\) coordinates in the image frame), we define multidimensional volatility of a sequence of gaze points \(X\) as Euclidean variance around the centroid

\[
v(X, t, \tau_{gaze}) = \frac{1}{\tau_{gaze}} \sum_{i=t}^{t-\tau_{gaze}} \| x_i - \mu(X_{t:t-\tau_{gaze}}) \|_2.
\]

We use this quantity to define a fixation when volatility \(v(X, t, \tau_{gaze})\) falls below a threshold which is updated to the 40th percentile of the volatility every 0.5 s. Figure 1 shows the gaze volatility (magenta) along with its threshold (yellow) and the selected gaze fixations (black circles). Based on results during preliminary analyses, we have set the values of the parameters to \(\tau_{rms} = \tau_{boll} = \tau_{gaze} = 300\) ms and \(\eta = 2\).

### B. Visual Feature Extraction

For each fixation, the gaze position will be used to obtain an image of the observed object from the first-person video recording of the scene. From this image, the object’s affordances will be encoded into appropriate visual features.

The video recording and the gaze tracking data are synchronized and expressed in the same reference system, thus the gaze point always lies on the object on which the user is focusing. We isolate this object from the others in the image using the Active Segmentation algorithm by Mishra et al. \[15\]. This method uses brightness, colors, and textures to segment the object on which the gaze falls. The drawback of the fixation-guided segmentation is its sensitivity to noise in the gaze position estimate. On the other end, its substantial advantage over object-detection methods based on machine
learning is that it does not require any prior knowledge about the appearance of the objects of interest.

The object’s affordances are extracted from its image and encoded into appropriate visual features using a Convolutional Neural Network (CNN) as a feature extractor. Deep visual features, indeed, are able to gather spatial and high-level visual characteristics, like shapes and color gradients. The object image is fed into the VGG-16 CNN pre-trained on ImageNet \[16\] and the activation of the second-last fully-connected layer is taken as the image visual feature.

Under the assumption that the object of interest remains the same during the whole prehension, the CNN feature associated with a certain fixation is maintained for all the subsequent samples, until the next fixation. A side effect of this choice is that each arm rest will be associated with the visual features of the object grasped in the previous prehension. However, this is inevitable because we do not know a priori the grasp’s duration.

C. Multimodal integration

At the end of the feature extraction process, the visual cue can be used alone or in conjunction with the myoelectric one to train a grasp classifier. Among the possible methods to integrate multimodal cues, we opt for mid-level integration \[17\], also known as integration at the kernel level. This method combines couples \((x, y)\) of multimodal samples of the type \(x = (x^1, \ldots, x^C)\) by computing their similarity via a weighted sum of cue-specific kernel functions

\[
k_{mc}(x, y) = \sum_{i=1}^{C} w_i k_i(x^i, y^i) \quad \text{for } w_i > 0 . \tag{4}
\]

The weights \(w_i\) of the kernel combination are free hyper-parameters of the multi-cue kernel. Such similarities will be used by a kernel machine classifier to create the prediction model.

IV. EXPERIMENTAL SETUP

We collected a custom dataset in which we recorded sEMG and gaze while subjects performed a set of grasps on different objects. In the following, we detail the dataset and how the data was used in our offline evaluation.

A. Dataset

Five intact subjects (4M, 1F) participated in our study. We selected ten grasps based on relevant literature \[18\] and on their perceived importance for Activities of Daily Living (ADL). Each of the grasps was performed on three representative objects that could reasonably be manipulated using the respective grasp. In selecting these objects, we attempted to re-use them as much as possible for multiple grasps to enforce a many-to-many relationship: grasps can be used with multiple objects and objects can be used with multiple grasps. This avoids the risk that an object’s identity alone is sufficient to unequivocally predict a grasp. During the acquisition, we made sure that there were always a minimum of five objects placed in front of the subject, to encourage realistic gaze behavior and to increase visual clutter.

Aside from multiple objects, the acquisition protocol was extended in two other manners to encourage variability in the myoelectric signals. One source of variability is given by the limb position effect, meaning that the signals will depend on the orientation of the limb. We took this into account by performing all movements both while seated and standing, which are likely the most common orientations in ADL. Second, we extended the protocol with either one or two functional tasks for each of the grasps. This introduces variability in the dynamic context of the hand, or more precisely crosstalk due to the added activity of muscles controlling the wrist and limb. Also in this case these functional movements were selected to represent ADL. The grasps, their respective objects, and the functional tasks are listed in Table I.

During the exercise, the subject was in front of a table on which the objects were placed. Prior to each grasp, a screen showed short movies of the movements with the aim to clarify how each of the objects should be approached. The scope of this video was to help the subject become familiar with the procedure and to perform a training trial while the video was playing. After this initial phase, the subject was requested to repeat each movement-object combination or functional movement four times. The computer indicated when to start the grasp and when to release via audio instructions. As visual support, the required grasp was schematically shown on the screen for the entire duration of the exercise. Each repetition took approximately 8 s, containing the actual grasp (4 s to 5 s) and the subsequent transition back to the rest posture (3 s to 4 s).

Muscular activity was recorded using twelve Delsys Trigno double differential sEMG electrodes placed in two rows around the forearm, where the upmost row contained eight electrodes and the remaining four were placed lower (see Figure 2). The myoelectric signals were sampled at 2 kHz. At the same time, the gaze and first-person scene video were recorded using the Tobii Pro Glasses II. These glasses record the subject’s gaze at 100 Hz with a theoretical accuracy and precision of 0.5° and 0.3° degrees RMS, respectively. The frame also contains a forward facing scene camera with a field of view of 90° that records Full HD video at 25 Hz. The onboard software of the Tobii glasses conveniently precomputes the gaze point in the reference frame of the gaze camera, which is what we will use in the remainder of the paper. Figure 2 gives an overview of the acquisition setup.

The acquisition laptop was used to assign timestamps to sEMG and gaze samples in a shared reference time. These timestamps were used during preprocessing to synchronize all modalities and to upsample them to the sampling rate of sEMG. Furthermore, we filtered powerline interference and corrected the labels using the relabeling method described by Gijsberts et al. \[19\].
| Grasps       | Objects | Task Description                  | Objects | Task Description                  |
|-------------|---------|-----------------------------------|---------|-----------------------------------|
| Medium Wrap |         | Take the Bottle while Seated/Standing |         | Drink from the Can while Standing |
|             |         | Take the Can while Seated/Standing |         | Open and close the Door Handle while Standing |
|             |         | Take the Door Handle while Seated/Standing |     |                                  |
| Lateral     |         | Take the Cup while Seated/Standing |         | Turn the Key in the lock while Standing |
|             |         | Take the Key while Seated/Standing |         | Open and close the Jacket while Standing |
|             |         | Take the Pencil Case while Seated/Standing |     |                                  |
| Parallel Extension |         | Take the Plate while Seated/Standing |         | Lift the Plate while Standing |
|             |         | Take the Book while Seated/Standing |         |                                  |
|             |         | Take the Drawer while Seated/Standing |     |                                  |
| Tripod Grasp |         | Take the Bottle while Seated/Standing |         | Open and close the cap of the Bottle while Standing |
|             |         | Take the Cup while Seated/Standing |         | Open and close the Drawer while Standing |
|             |         | Take the Drawer while Seated/Standing |     |                                  |
| Power Sphere |         | Take the Ball while Seated/Standing |         | Move the Ball to the right and back while Standing |
|             |         | Take the Light Bulb while Seated/Standing |     |                                  |
|             |         | Take the Key while Seated/Standing |         |                                  |
| Precision Disk |       | Take the Jam Jar while Seated/Standing |         | Open close the lid of Jam Jar while Seated |
|             |         | Take the Light Bulb while Seated/Standing |     | Screw and unscrew the Light Bulb while Seated |
|             |         | Take the Ball while Seated/Standing |         |                                  |
| Prismatic Pinch |      | Take the Clothespin while Seated/Standing |         | Squeeze the Clothespin while Seated |
|             |         | Take the Key while Seated/Standing |         |                                  |
|             |         | Take the Can while Seated/Standing |         |                                  |
| Index Finger Extension |     | Take the Remote Control while Seated/Standing |         | Press a button on the Remote Control while Seated |
|             |         | Take the Knife while Seated/Standing |         | Cut bread with the Knife while Seated |
|             |         | Take the Fork while Seated/Standing |         |                                  |
| Adducted Thumb |       | Take the Screwdriver while Seated/Standing |         | Turn the Screwdriver while Seated |
|             |         | Take the Remote Control while Seated/Standing |     |                                  |
|             |         | Take the Wrench while Seated/Standing |         |                                  |
| Prismatic Four Fingers |      | Take the Knife while Seated/Standing |         | Move the Fork to the right and back while Seated |
|             |         | Take the Fork while Seated/Standing |         |                                  |
|             |         | Take the Wrench while Seated/Standing |         |                                  |
The four repetitions of a movement is used once to test the training/test splits of the database, such that each of the four splits are used once to test the model while the remaining three are used as training set. Subsequently, the prediction accuracy is averaged over the four splits. For computational reasons, we subsampled the test data at a factor 20, meaning that effectively we predict a sliding window with interval of 10 ms. The training data instead was subsampled with an additional factor of 10, while the data used for hyperparameter optimization was subsampled with a factor of 10 · 4. Besides our multimodal classifier, we also include single cue classifiers as reference and a baseline that predicts simply the most common class in the training data. In our specific case, this means predicting always the “rest” class, since this is the most common class due to our acquisition protocol.

V. RESULTS

The goal of this section is to determine if the standard sEMG approach would benefit from the integration of the visual cues found by our algorithm. Figure 3 reports the average classification accuracy of the four classifiers for each subject. The sole visual cue does not produce a considerable improvement in accuracy with respect to the baseline, as the average improvement is of the 2% and it is mainly due to two of the five subjects. Nevertheless, the integration of vision to the muscular cue increases the average accuracy of more than 4% over that of the EMG classifier, and this result confirms our initial guess that the visual cue conveys complementary information with respect to the muscular one and that their integration improves the performance of the grasp classification task.

The contribution of each of the two cues to the multimodal classification is indicated by the values of the weights wEMG and wCNN that the algorithm automatically choose (during hyperparameter optimization) to combine the cues at the kernel level. Figure 4 reports the values of the kernel weights.
for each subject and demonstrates how the contribution of the two cues is balanced, being around the 65% for the muscular cue and the 35% for the visual one on average.

We also analyze the distribution of the prediction error during the different phases of the prehension. Each prehension of the experiment has a different duration after relabeling, but always consists of a grasp preceded by a rest phase. In Figure 5, we report the prediction error with respect to the normalized duration of the rest phase (−1, 0) and subsequent grasp ([0, 1]). The addition of visual cues consistently reduces the prediction error during the grasp (t ∈ [0, 1]). Relevantly, the most consistent reduction in prediction error due to the visual cue happens at the onset and at the offset of the grasp. This indicates that vision compensates for the increased level of noise in the myoelectric signals during movement transitions. At the same time, the visual cue causes a slightly higher prediction error during the rest phase. This is because, generally, the visual information related to arm rest comes from the previous grasp and is, in fact, misleading for the classification of the “rest” movement. As already explained in subsection III-B, this side-effect of the visual feature propagation is inevitable because we do not know in advance the duration of the grasps.

Qualitatively, the classification improvement obtained by integrating CNN and EMG can be observed by subtracting the confusion matrix of EMG+CNN to that of EMG. This difference is shown in Figure 6. The positive values on the diagonal indicate a uniform improvement of the classification accuracy for all the 10 grasps. However, the negative value at location (1,1) indicates an increase in rest misclassification and confirms the considerations made about the effect of holding the visual cues also during rest.

In tasks where the classification predictions form a temporal sequence, it is advisable to define performance metrics that distinguish if errors are caused by misclassifications or delays in the predictions. This is useful, for instance, to observe the effects of smoothing the series of predicted movements via a majority vote. When the dimension k of the majority vote window is increased, the number of misclassifications generally decreases at the expense of a higher prediction delay. Standard classification accuracy fails to catch these competing effects, hence we will analyze the classifier performances also using the Movement Error Rate (MER) and the prediction delay, proposed by Gijsberts et al. [19]. The MER measures the similarity between the true and the predicted series of movements rather than considering the accuracy of the classification sample by sample. This quantity is insensitive to delays in the prediction, which are instead measured via the prediction delay, defined as the average time interval between a label change and the first correct prediction. Figure 7 represents the values of MER and delay achieved by the EMG (blue) and the EMG+CNN (red) classifiers when varying the length k of a majority vote window (k ∈ {1, 3, 5, 11, 25, 50, 100, 150, 250}). The integration of visual features to muscular ones proves to reduce the MER consistently for all the considered values of k. In particular, for k < 50, EMG+CNN shows a halved MER with an unchanged prediction delay. This shows that the reduction in error of the EMG+CNN classifier does not come at the cost of increased delay.

VI. CONCLUSIONS

This work demonstrated how standard sEMG based grasp classification benefits from the integration of the affordances of the manipulated objects. We proposed a method to automatically extract the object affordances from a first-person video recording of the scene and an estimate of the gaze position. The method identifies relevant gaze fixations on the
base of ocular and muscular activity. The objects observed during such fixations are segmented and their affordances are encoded into high-level visual features, extracted by an off-the-shelf Convolutional Neural Network. Despite we only conducted an offline evaluation of the method, the fixation detection has been designed to follow an online execution paradigm.

The method was evaluated on the data collected from intact subjects performing several of the most common grasps in activities of daily living. The acquisition protocol has been designed to simulate the prosthesis usage in a realistic environment. To ensure variability, we considered grasps both in a static setting as well as when used to perform a functional task, while we took the limb position effect into account by repeating the movements while seated and standing. Furthermore, the same objects were associated to multiple grasps to enforce a many-to-many relationship between grasps and objects, and multiple objects were placed in the user’s field of view to encourage realistic gaze behavior.

Our tests confirmed that the integration of object affordances to the muscular activity of the forearm is indeed useful for grasp classification. The average prediction accuracy went from 80%, when using only the EMG cue, to 84%, when integrating EMG and vision. This improvement was considerable, as it involved uniformly all the subjects and all the grasp types. As expected, the contribution of vision was higher at the onset and the offset of the grasp, when the myoelectric cue is affected by motion artifacts. Finally, the analysis of the Movement Error Rate suggested that the performances of the multimodal classifier can be further reduced with a majority vote of the predictions at no expense of the prediction delay.

REFERENCES

[1] B. Peerdeman, D. Boere, H. Witteveen, R. H. 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,” Journal of Rehabilitation Research and Development, vol. 48, no. 6, pp. 719–738, 2011.
[2] A. Fougner, E. Scheme, A. D. C. Chan, K. Englehart, and Ø. Stavdahl, “A multi-modal approach for hand motion classification using surface emg and accelerometers,” in Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2011, pp. 4247–4250.
[3] S. Đošen, C. Cipriani, M. Kostić, M. Controzzi, M. C. Carrozza, and D. B. Popović, “Cognitive vision system for control of dexterous prosthetic hands: Experimental evaluation,” Journal of NeuroEngineering and Rehabilitation, vol. 7, no. 1, p. 42, 2010. [Online]. Available: http://dx.doi.org/10.1186/1743-0003-7-42
[4] S. Đošen and D. B. Popović, “Transradial prosthesis: Artificial vision for control of prehension,” Artificial Organs, vol. 35, no. 1, pp. 37–48, 2011. [Online]. Available: http://dx.doi.org/10.1111/j.1525-1594.2010.01040.x
[5] Y. Fang, Nalinda, D. Zhou, and H. Liu, “Multi-modal sensing techniques for interfacing hand prostheses: A review,” IEEE Sensors Journal, vol. 15, no. 11, pp. 6065–6076, Nov 2015.
[6] J. Lobo-Prat, P. N. Kooren, A. H. Stienen, J. L. Herder, B. F. Koopman, and P. H. Veltink, “Non-invasive control interfaces for intention detection in active movement-assistive devices,” Journal of NeuroEngineering and Rehabilitation, vol. 11, no. 1, p. 168, 2014. [Online]. Available: http://dx.doi.org/10.1186/1743-0003-11-168
[7] A. Gijsberts and B. Caputo, “Exploiting accelerometers to improve movement classification for prosthetics,” in 2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR), June 2013, pp. 1–5.
[8] Y. Hao, M. Controzzi, C. Cipriani, D. B. Popović, X. Yang, W. Chen, X. Zheng, and M. C. Carrozza, “Controlling hand-assistive devices: utilizing electrooculography as a substitute for vision,” IEEE Robotics Automation Magazine, vol. 20, no. 1, pp. 40–52, March 2013.
[9] M. Markovic, S. Đošen, C. Cipriani, D. Popović, and D. Farina, “Stereo vision and augmented reality for closed-loop control of grasping in hand prostheses,” Journal of
[10] M. Marković, S. Došen, D. Popović, B. Graimann, and D. Farina, “Sensor fusion and computer vision for context-aware control of a multi degree-of-freedom prosthesis,” *Journal of Neural Engineering*, vol. 12, no. 6, p. 066022, 2015. [Online]. Available: http://stacks.iop.org/1741-2552/12/i=6/a=066022

[11] G. Ghazaei, A. Alameer, P. Degenaar, M. Gorgan, and K. Nazarpour, “An exploratory study on the use of convolutional neural networks for object grasp classification,” in *2nd IET International Conference on Intelligent Signal Processing 2015 (ISP)*, Dec 2015, pp. 1–5.

[12] E. A. Corbett, N. A. Sachs, K. P. Körding, and E. J. Perreault, “Multimodal decoding and congruent sensory information enhance reaching performance in subjects with cervical spinal cord injury,” *Frontiers in Neuroscience*, vol. 8, p. 123, 2014. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fnins.2014.00123

[13] D. P. McMullen, G. Hotson, K. D. Katyal, B. A. Wester, M. S. Fifer, T. G. Mcgee, A. Harris, M. S. Johannes, R. J. Vogelstein, A. D. Ravitz, W. S. Anderson, N. V. Thakor, and N. E. Crone, “Demonstration of a semi-autonomous hybrid brain-machine interface using human intracranial eeg, eye tracking, and computer vision to control a robotic upper limb prosthesis,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 4, pp. 784–796, July 2014.

[14] R. S. Johansson, G. Westling, A. Bäckström, and J. R. Flanagan, “Eye–hand coordination in object manipulation,” *Journal of Neuroscience*, vol. 21, no. 17, pp. 6917–6932, 2001.

[15] A. K. Mishra, Y. Aloimonos, L.-F. Cheong, and A. A. Kassim, “Active visual segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 639–653, April 2012.

[16] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556

[17] T. Tommasi, F. Orabona, and B. Caputo, “Discriminative cue integration for medical image annotation,” *Pattern Recognition Letters*, vol. 29, no. 15, pp. 1996–2002, 2008.

[18] T. Feix, J. Romero, H. B. Schmiedmayer, A. M. Dollar, and D. Kragic, “The grasp taxonomy of human grasp types,” *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 66–77, Feb 2016.

[19] A. Gijsberts, M. Atzori, C. Castellini, H. Müller, and B. Caputo, “Movement error rate for evaluation of machine learning methods for semg-based hand movement classification,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 4, pp. 735–744, July 2014.

[20] R. Rifkin, G. Yeo, and T. Poggio, “Regularized least squares classification,” in *Advances in Learning Theory: Methods, Model and Applications*, vol. 190. VIOS Press, 2003, pp. 131–154.