IMPROVING OPTICAL MYOGRAPHY VIA CONVOLUTIONAL NEURAL NETWORKS

Christian Nissler*, Imran M. Badshah*, Claudio Castellini*, Wadim Kehl² and Nassir Navab²

**Institute of Robotics and Mechatronics, German Aerospace Center (DLR)*
²Institut für Informatik,Technische Universität München (TUM)

ABSTRACT

In order to improve the accuracy and reliability of myocontrol (control of prosthetic devices using signals gathered from the human body), novel kinds of sensors able to detect muscular activity are being explored. In particular, Optical Myography (OMG) consists of optically tracking and decoding the deformations happening at the surface of the body whenever muscles are activated. OMG potentially requires no devices to be worn, but since it is an advanced problem of computer vision, it incurs a number of other drawbacks, e.g., changing illumination, identification of markers, frame tear and drop. In this work we propose an improvement to OMG as it has been recently introduced, namely we relax the need of precise positioning and orientation of the markers on the body surface. The small size of the markers and their curvature while adhering to the surface of the forearm can lead to missed detections and misdetections in their orientation; here we rather detect the deformations by applying a Convolutional Neural Network to the region of interest around the feature source segmented, from the forearm. The classification-based approach yields results similar to those obtained by other classification based modalities, reaching accuracies in the range of 96.21% to 99.30% when performed on 10 intact subjects.

INTRODUCTION

Recently introduced in the scientific community of assistive robotics, Optical Myography (OMG) is a novel way of non-invasively detecting the muscular activity [1]. The main idea is simple: muscle activation, for instance when flexing the index finger, induces a quite precise deformation in the forearm due to the enlargement of the muscle belly, and the consequent shifting of the adjacent musculoskeletal structures. This phenomenon is currently being exploited by the techniques called Force Myography (FMG) and Tactile Myography (TMG); in these cases, the aforementioned deformations are detected via pressure / force sensors, used in small numbers (FMG) or in a high-resolution array (TMG). The results are highly promising [2] [3]. In OMG, such deformations are captured using optical recognition alone, that is, by "looking" at the forearm. Whenever the fingers flex, or the wrist rotates and/or extends, small changes in the forearm's volume and shape appear to the trained eye and become apparent if markers are applied to the surface of the forearm; they can be linearly related to the required muscle activations [1] [4]. This possibility leads to applications in, e.g., advanced upper-limb prosthetics: a camera aimed at the stump of an amputee could be able to control the device, either in the real world or in an Augmented / Virtual Reality setup. The main advantage of OMG is probably that it is the "ultimate" non-invasive human-machine interface for the disabled, since it requires no equipment to be placed on the forearm / stump (in principle even markers can be avoided, if the algorithm is smart enough). On the other hand, it naturally suffers of the problems commonly associated to computer vision: dependence on the illumination and focus, loss of precision due to the varying distance between camera and subject, occlusions, missed and/or misdetection of the markers, and surface features of the forearm. In this work we propose an advancement to [1], in which AprilTags [4] [5] were used and tracked by a camera during movement of the fingers. In a new experiment, we rather used a plain sticker, whose deformations was observed by a single camera and then passed to a Convolutional Neural Network (CNN). With this approach, the need for a fixed relation of the camera to the arm is mitigated and the camera can be worn at the arm. Our experimental results show that the classification of five different finger poses are in the range of 96.21% to 99.3% and therefore on par with state of the art methods like surface EMG, Ultrasound or Force Myography.

Related work

Muscle activity exists even after the amputation of the hand [6] [7]. Such activity can be non-invasively detected in a number of ways: through the electrical activity of the motor units (surface electromyography or sEMG [8]), the deformations of the involved body parts (force or tactile myography [2]), listening to the vibrations induced by the muscle motion (mechanomyography [9]), and so on. We hereby concentrate on another modality, termed Optical Myography (OMG), which estimates finger poses by optically observing deformations on the surface of the forearm. This is performed by mounting a web-camera to a fixed set-up frame and by strapping the subject's forearm to it. By preventing the forearm to move relative to the camera, solely its muscle deformations can be detected with the help of fiducial markers such as AprilTags [4]. The 6-D information (translation and rotation) collected from each of these tags are processed and regressed to four different finger poses (trained independent of one another). The small size of the markers can lead to its improper identification. The curvature when stuck to the surface of the forearm also leads to misdetections in orientation. These factors magnify once the arm is released from the set-up in attempt to carry out practical tasks.

EXPERIMENT DESCRIPTION

Setup Description

An elastic band is used to strap a camera onto the forearm. The images are recorded at a resolution of 640x480 pixels at a frame rate of 25 frames per second. Artificial lighting (from LED lights) is used in these experiments to achieve uniformity in the experiments and to avoid changes in illumination caused by natural light. Motion blur was also physically suppressed to a certain extent by using a velcro band around the forearm and the camera to prevent upwards vertical movement. The images are pre-processed to obtain the ROI using the computer vision library OpenCV and then passed to a CNN, which is implemented using the TensorFlow software library.

Figure 1: The experiment setup, where a subject is following the stimuli on the monitor with a sticker stuck on to the left forearm and a camera attached to the arm to capture its deformations

Participants

The participants chosen for the experiment were people with all their fingers intact. Each subject was asked to use around 80% of their maximum force while following a stimulus signal displayed by a virtual hand on a computer screen. The camera was strapped on the subject's arm to simulate an attachment to the base of an active hand prosthesis as shown in Figure 2.

There were two subjects with the dominant hand being their left, while the others were right handed. The average age of the participants (three female and 7 male) is 26.2 ±3.65 years. A plain sticker was stuck on to the anterior side of each subject's arm. Once prepared for the experiment, the subject was asked to place the forearm (freely) on a plain

Figure 2: (left) The sticker attached to the forearm of a subject; (right) segmentation of the sticker

surface. The subject was then shown both the graphical interface used for recording and a virtual 3-D hand model presenting the stimulus signal to be followed. The stimuli used are thumb flexion, thumb abduction (rotation), index flexion, combo flexion (combination of the little, ring and middle finger) and rest (all fingers relaxed), repeated 10 times at equal intervals. The experiments were approved by the Ethical Committee of the DLR and all subjects gave written consent.

Image processing

By using a plain sticker on the forearm, the sticker's bounding box can be used as a region of interest (ROI) and its deformations and slight changes in position can be used as characteristics for the CNN to distinguish between the finger poses. This requires a segmentation of the sticker in each camera image from the background. In order to segment the sticker, the RGB colour space is transformed into a three channel log opponent chromacity (LO) space, a method common in skin segmentation algorithms [10] [11]. In order to cope with intensity issues such as glare, the intensity channel (I channel) is subtracted from the LO-R_g channel and this new channel is used as the base for the rest of the image processing upon normalization.

A fixed ROI is set for the first frame in order to remove unnecessary background. A median filter smoothens the image before an adaptive threshold is used to segment the boundaries. Morphological operations are used in cases where the sticker's boundary merges with the forearm's or the sleeve's due to its placement. The contours of the sticker are then extracted and enclosed within a rectangular bounding box. Necessary conditions are imposed on the bounding box to filter the contours of the sticker. Once the estimated contour and bounding box of the sticker is verified by the user in the first frame (the remaining being automated), a mask of the contour segments only the sticker and sets the regions outside the contour to a pixel value of 0 (black).

Classification method

The network is a simple CNN consisting of two convolution layers with exponential linear unit (ELU) activation and a fully connected layer. Each convolution layer consists of 16 filters. The filters in the first layer have a relatively large size of 11x11 pixels while the second layer's filters are of size 5x5 pixels. The images sent as input to the CNN are greyscale images of size 130x130 pixels. Image sizes are restricted to be smaller than the down-scaled image to preserve information. About nearly a second (800 ms) of delay is imposed to adapt for the reaction time of the subject. Intermediate pose data are also not considered as an input in order to make it a purely classification problem. The data sent in for training are shuffled in a pseudorandom process so that each of the stimuli is uniformly distributed throughout the training set. The training set is then split into batches of 70. The predictions are estimated by optimizing the parameters of the CNN using stochastic gradient descent which minimizes the loss over 20 epochs. The initial learning rate is 0.001; which is then decreased by 5% every succeeding epoch. The loss function is that of the mean of the sparse softmax cross-entropy of the output of the final fully connected layer (the logits). The accuracy is calculated by finding the argmax of the logits and comparing it to the true labels.

"Leave-one-repetition-out" cross-validation was used to evaluate subject accuracy, that is, for each subject and repetition we trained the CNN on nine repetitions and tested on the selected one. We averaged out all results across subjects to yield a global result.

RESULTS

Figure 3 shows the overall accuracy, i.e. the percentage of correctly classified cases versus all observed cases, averaged over all subjects as well as overall precision, i.e. the true positives versus all positives. The mean and standard deviation cannot be taken as the best value over the repetitions since the overall (intra-subject) sample space is small (10 repetitions yielding 10 separate tests). Thus the median and interquartile range (IQR) depict an evaluation closer to that of the true performance of the model. The confusion matrix of the global median is then used to obtain the global accuracy and precision. The final results are displayed in percentage (after normalization).

As a comparison with existing literature, we show in Figure 4 that OMG using CNN (OMG_CNN) performs on par with

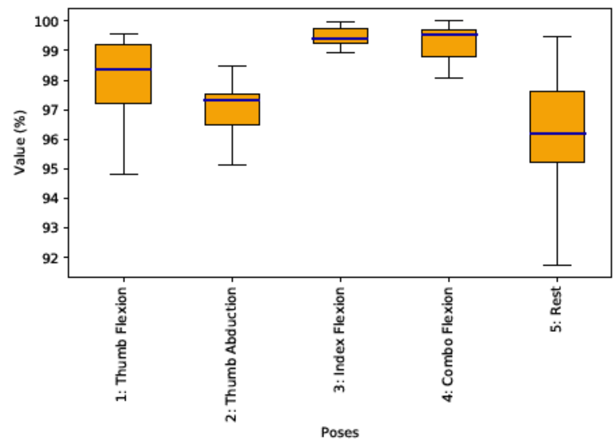

Figure 3: Boxplot of classification accuracy of each pose

the other classification-based modalities. We use the results obtained during similar experiments published in the following papers:

Naik et al. [12], Table II, III(A) and III(B) (labelled sEMG in our Figures): Five transradial amputees were engaged in performing 11 finger poses, which were detected using two proposed sEMG configuration which they considered optimal. We take the average of the little, ring, middle, pointer and thumb extension, and the little, ring and middle finger flexion as the rest pose and combo flexion respectively. The remaining poses were the same as in this work.

Cho et al. [2], Table 2 and Figure A1 (FMG): Four transradial amputees performed five trials with 11 different grip gestures. The key grip performed in their study was assumed similar to the thumb flexion in our work, the mouse grip was almost the same as the thumb abduction, the precision open is assumed similar to the index flexion and the finger point and relaxed hand were the same as the combo flexion and rest position respectively. The final confusion matrix drawn from the four subjects' was that of their mean as used by the OMG method between the subjects.

Sikdar et al. [13], Table I (SMG): Ten healthy volunteers performed individual finger flexions. The combo flexion was derived by taking the average of the performance by the little, ring and middle finger flexion, while the remaining were the same as in this work. The rest and the thumb abduction could not be gathered for the comparison.

Consider the bar plots in Figure 4. The index and combo flexion are in fact as high as achieved when using sEMG. One can note that the thumb flexion, thumb abduction and rest position has a lower precision (Figure 4b). This can be explained by the low variation between the three poses, which causes a higher risk of false positives.

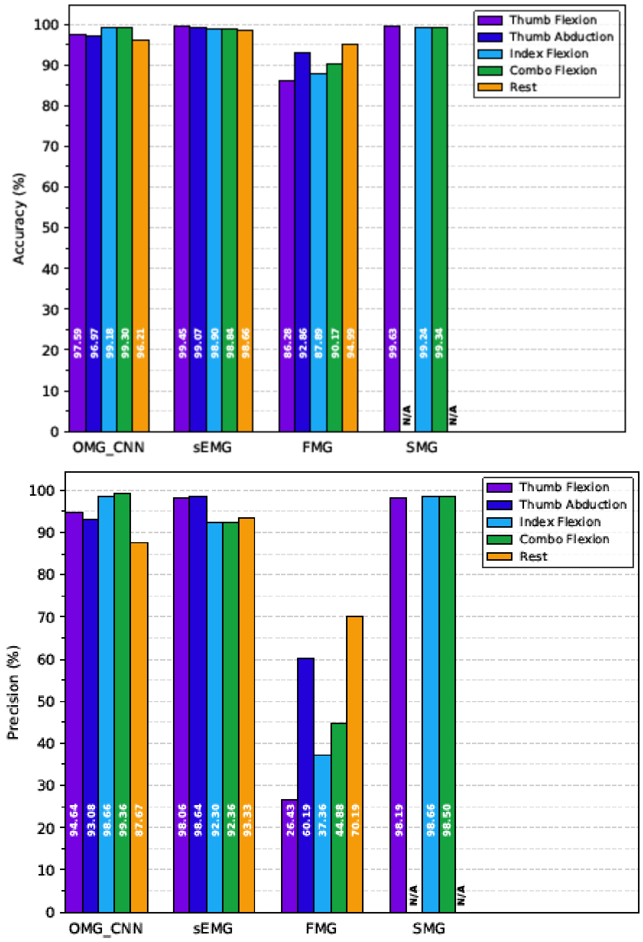

Figure 4: Modality-wise comparison of accuracy (top) and precision (bottom) in percentage between the various finger poses

CONCLUSIONS

In this paper we have proposed a simple improvement to optical myography (OMG), namely, the usage of a single undifferentiated marker ("sticker") instead of several AprilTags, as it was done in [1]. We have demonstrated that such a simple arrangement is enough to obtain classification results, for several movements of the fingers, which are comparable to those already obtained in literature using sEMG, ultrasound imaging and force myography. A Convolutional Neural Network seems to be a good option to take advantage of the image-like nature of the sticker and its deformation due to muscular activity. The main challenges to be faced here are the morphological operations used to disjoin boundaries intersecting with the sticker's during contour extraction. It is interesting to note that thumb movements are not easily distinguishable from each other and from the rest pose, probably due to the fact that the muscles, which control the thumb are deeper and therefore harder to detect on the skin surface using an optical camera.

OMG is, of course, susceptible to all well-known pitfalls of computer vision: motion blur, varying illumination and occlusion(s). This is going to be the main line of future research, especially as we will lift the assumption of the forearm being fixed in a specific spatial position.

REFERENCES

[1] C. Nissler, N. Mouriki, and C. Castellini, “Optical Myography: Detecting Finger Movements by Looking at the Forearm,” Frontiers in Neurorobotics, vol. 10, no. 3, 2016.

[2] E. Cho, R. Chen, L.-K. Merhi, Z. Xiao, B. Pousett, and C. Menon, “Force myography to control robotic upper extremity prostheses: a feasibility study,” Frontiers in bioengineering and biotechnology, vol. 4, 2016.

[3] A. Radmand, E. Scheme, and K. Englehart, “High-density force myography: A possible alternative for upper-limb prosthetic control” J Rehabil Res Dev., vol. 53, no. 4, pp. 443–456, 2016.

[4] C. Nissler, N. Mouriki, C. Castellini, V. Belagiannis, and N. Navab,“OMG: Introducing optical myography as a new human machine interface for hand amputees,”Proceedings of ICORR – International Conference on Rehabilitation Robotics, 2015, pp. 937–942.

[5] E. Olson, “Apriltag: A robust and flexible visual fiducial system” 2011 IEEE International Conference on Robotics and Automation(ICRA). IEEE, 2011, pp. 3400–3407.

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

[7] B. Peerdeman, D. Boere, H. Witteveen, H. Hermens, S. Stramigioli, J. Rietman, P. Veltink, S. Misra, et al., “Myoelectric forearm prostheses: State of the art from a user-centered perspective,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2011.

[8] 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.

[9] M. A. Islam, K. Sundaraj, R. B. Ahmad, and N. U. Ahamed, “Mechanomyogram for muscle function assessment: A review,” PLoS ONE, vol. 8, no. 3, p. e58902, 03 2013.

[10] W. R. Tan, C. S. Chan, P. Yogarajah, and J. Condell, “A fusion approach for efficient human skin detection,” IEEE Transactions on Industrial Informatics, vol. 8, no. 1, pp. 138–147, 2012.

[11] T. A. El-Hafeez, “A new system for extracting and detecting skin color regions from pdf documents,” International Journal on Computer Science and Engineering, vol. 2, no. 9, pp. 2838–2846, 2010.

[12] G. Naik, A. Al-Timemy, and H. Nguyen, “Transradial amputee gesture classification using an optimal number of semg sensors: an approach using ica clustering,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2015.

[13] 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,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 22, no. 1, pp. 69–76, 2014.