Combining Electromyography and Fiducial Marker Based Tracking for Intuitive Telemanipulation with a Robot Arm Hand System

Anany Dwivedi, Gal Gorjup, Yongje Kwon, and Minas Liarokapis

Abstract—Teleoperation and telemanipulation have since the early years of robotics found use in a wide range of applications, including exploration, maintenance, and response in remote or hazardous environments, healthcare, and education settings. As the capabilities of robot manipulators grow, so does the control complexity and the remote execution of intricate manipulation tasks still remains challenging for the user. This paper proposes an intuitive telemanipulation framework based on electromyography (EMG) and fiducial marker based tracking that can be used with a dexterous robot arm hand system. The EMG subsystem captures the myoelectric activations of the user during the execution of specific hand postures and gestures and translates them into the desired grasp type for the robot hand. The pose of the tracked fiducial marker is used as a task-space goal for the robot end-effector. The system performance is experimentally validated in a remote operation setting, where the system successfully performs a telemanipulation task.

I. INTRODUCTION

Robot systems have over the past decades seen great technological development, bringing them to remarkable levels of speed and accuracy and making them invaluable for tasks that require precise repetitive motions. These systems can either be programmed to work autonomously, or guided by human operators in cases when autonomous operation is infeasible or undesired. In such teleoperation frameworks, a remote user can have partial or full access to the robot that accomplishes their intent. With human support, robot systems are able to perform complex tasks that would have been infeasible to execute autonomously.

One of the key teleoperation applications is task execution using human skill and expertise in remote or extreme environments. For example, in NASA’s Robonaut project [1], a teleoperated robot was developed for space station maintenance and deep space exploration missions. During the nuclear leak in the Fukushima Daiichi power plant in 2011, the plant surroundings were deemed too dangerous for maintenance through direct human involvement and remotely operated robots were considered for the mission [2]. Another instance are remotely operated underwater vehicles (ROVs) that are used in deep water inspection and object retrieval missions, such as the inspection of the Titanic [3] or the retrieval of hydrogen bombs lost in the Mediterranean sea [4]. ROVs are also used for underwater maintenance missions like sealing crevices in the hull of the Prestige oil tanker [5]. In addition to applications in hostile environments, teleoperated systems also find a wide range of use cases in healthcare, especially surgery, due to two main reasons. The first is augmenting the surgeon’s capabilities by increasing the range of motion of the surgical tool and allowing them to exert forces with great precision. The second is reducing travel related time delays by allowing the surgeon to apply their expertise without spending time on transport [6], [7]. Other teleoperation applications can be found in the field of research and education. With the availability of remote labs and hardware platforms, researchers and students can test their algorithms on physical systems, making it easier for them to learn and understand fundamental concepts [8].

With the increasing complexity of teleoperated devices, hand-held controllers become inadequate for intuitive and non-fatiguing teleoperation [9]. Complicated interfaces in such schemes may confuse the user and impede their motion, which raises a need for a more natural means of controlling the robot. One of the most intuitive human-machine interfacing options is electromyography (EMG), which allows the user to employ natural arm movement and gestures to control a robot arm hand system [10]. Moreover, EMG based control allows the user to perform other tasks while interacting with the teleoperated system as their hands are kept free [11].

This paper presents an affordable and intuitive telemanipulation framework that can be used with a dexterous robot arm hand system. All components are mounted on the user, making the system highly portable. The proposed framework utilizes EMG and fiducial marker based pose tracking to allow natural control of the robot without constraining the hand movement (Fig. 1). The framework is easy to set up without any room preparation. The rest of the paper is organised as follows: Section II discusses the related work, Section III describes the framework and the experimental setup, Section IV discusses the obtained results, while Section V concludes the work and discusses future directions.
II. RELATED WORK

Since its advent in mid-20th century, the field of teleoperation and telemanipulation has seen significant development as one of the core aspects of robotics. In its early stages, remote manipulation with arm hand systems was explored mostly from a control theoretic perspective [12], [13]. Interfaces generally took the form of classic I/O devices (joysticks, keyboards, and monitors) or master/slave robots sharing a similar kinematic structure to provide intuitive control to the user. With time, the robot system functionality expanded and alternative interfacing options were explored to maximise control efficiency and ease of use.

A common solution for remote robot arm control that is intuitive for the user is mapping hand motion to the tool position and orientation. Several motion capture options are available for tracking the hand pose, including systems based on inertial [14] or magnetic [15] motion sensors mounted on the user. Markerless systems for human motion tracking have also been explored, but the data obtained through vision systems are noisy and not as reliable as the alternatives [16]. Some works focused on controlling robot arms without explicit pose information, but instead through human muscle activation interpretation. In [10], for instance, the authors proposed a methodology utilising low-dimensional representations of human myoelectric activations for controlling the motion of a robot arm. In [17], authors presented a machine learning system that mapped muscle activity to hand position, orientation, and grasping force, allowing them to control a robot arm in real-time. With the development of Virtual Reality (VR) technology, the introduced interfaces were promptly considered for telemanipulation [18]. Recent examples include using the HTC Vive VR system to collect teleoperation data in a learning framework for a PR2 robot [19]. In [20], the authors utilised the Leap Motion tracking system to obtain hand and finger poses and used them in a gesture-based, VR powered teleoperation framework. Many such systems come at a considerable cost related to the quality of captured motion data. Motion tracking based on fiducial markers is an affordable solution that requires only a regular camera for operation and offers pose estimates of reasonable quality. Their use in teleoperation was initially considered in [21], but they have not yet been effectively used for capturing target pose data for robot arm manipulators. However they have been utilised for enhancing the teleoperation environment [22] and tracking [23].

The complexity of end-effector control depends on the type of gripper or robot hand used in the framework. Trivial parallel grippers, for example, can be managed by a simple button or switch accessible by the operator. For dexterous, possibly anthropomorphic robot hands, a popular choice is data gloves, which may track finger poses, forces and wrist orientation [24], [25]. Selected VR interfaces mentioned above allow alternative means of finger pose tracking or offer inputs in the form of buttons and trackpads. An intuitive approach for controlling robot hands, particularly those with some degree of anthropomorphism, is through EMG. The authors of [26], for example, used a combination of autoregressive modelling and artificial neural networks to distinguish between different grasp types. In [27], the authors used myoelectric activations from flexor digitorum superficialis and flexor carpi ulnaris to manipulate a geometric computer model of the wrist and finger joints.

III. METHODS

A. Framework Structure

An overview of the telemanipulation framework structure is presented in Fig. 2, where the two main component groups are highlighted. Each block layout conceptually corresponds to a node architecture implemented within the Robot Operating System (ROS) [28], which provided the necessary communication, testing and visualisation utilities.

The Grasp Identification group focuses on decoding the user’s intent (triggering of a telemanipulation action and grasp type selection) and interpreting it for the intuitive control of the New Dexterity adaptive, humanoid robot hand [29]. The user muscle activations are captured by the EMG signal acquisition system and passed to the Intention Decoding module which produces posture class labels (presented in more detail in Section III-B) that correspond to specific grasp types. The resulting grasp types are then passed to the robot hand controller that handles the low level motor control. A selected hand posture is used as an enable signal for activating the telemanipulation framework.
The Arm Trajectory Generation block maps the user’s motion to robot arm (Universal Robots UR5) motion. The video stream captured by the head-mounted camera is passed to an image processing module that extracts the fiducial marker pose with respect to the camera frame (further described in Section III-C). The obtained marker pose is passed to the Pose Mapping and Inverse Kinematics module, which maps it to the robot end-effector and computes the corresponding inverse kinematics, obtaining the target robot joint angles. The goal robot configuration $q$ is forwarded to the Robot Arm Controller, which is based on the UR modern driver [30] and ROS Control [31] packages. To enable smooth and responsive real-time pose tracking, a closed-loop velocity controller for the UR robots was implemented because the joint speed commands provide best performance with the used robot models and control boxes [32]. The controller loops at 125 Hz, corresponding to the maximum frequency allowed by the UR system real-time interface.

B. EMG Based Intent Decoding

For the EMG based intent decoding, the classification models were trained offline on the acquired data. The myoelectric signals were recorded using surface EMG electrodes (DIN EMG snap cables attached to EMG stickers). The acquired signals were preprocessed by an EMG bioamplifier (g.Tec g.USBamp). The sampling rate for data acquisition was 1200 Hz. A Butterworth bandpass filter was applied to each channel (5Hz high-pass, 500Hz low-pass) with a 50Hz notch filter to reduce the line noise (see Fig 3).

For the experiments, each subject was instructed to alternate between rest state and a posture state (pinch grasp, tripod grasp, power grasp, and a co-contraction of all muscles that is used for enabling the telemanipulation scheme) on the presentation of an audio cue. The audio cues were presented using National Instruments (NI) MyRio. The MyRio was also used to trigger the bioamplifier when the audio cue was presented to isolate the rest phase and gesture phase during the training of the classification model. For each of the gestures, 5 trials were recorded with 20 repetitions each. The myoelectric activations of 8 muscle groups of the hand and forearm were recorded using double-differential electrodes. More precisely, the first channel was placed at the back of the hand focusing on the second dorsoral interossei to capture the flexion motion. The second channel was placed on the opponens pollicis to capture the activity of the thumb. Channels 3, 4, and 5 were placed on the flexor digitorum muscle site. Channels 6 and 7 were placed on the extensor digitorum muscle site and channel 8 was placed on the bicep brachii muscle [33], [34]. The ground electrode was placed on the elbow. The electrode placement is depicted in Fig. 4.

After the data was acquired, it was rectified and smoothed with a window of 20 samples before training the classifier (Fig. 3). Three different classification algorithms were used to discriminate between the examined gestures based on the acquired EMG data: 1) a Linear Discriminant Analysis (LDA) classifier, 2) a Random Forest (RF) classifier (an ensemble classifier based on decision trees), and 3) a Support Vector Machine (SVM) classifier. The classifiers were trained and tested using the 5-fold cross validation method.
C. Fiducial Marker Based Pose Tracking

The real-time pose tracking component of the telemanipulation framework was based on the ArUco class of fiducial markers [35]. The markers were attached to the user’s arm and captured by a commercial Logitech Brio webcam mounted on their head, streaming 720p video at 60 Hz. To compensate for the subject head motion and consequent shifting of the view angle, a reference marker was attached on the user’s shoulder. The marker detection and pose estimation was realized using the OpenCV [36] implementation of the ArUco Library, which extracted 6-DoF marker poses with respect to the camera frame. An example video frame with overlaid marker edges and corresponding pose estimates is presented in Fig. 5.

The homogeneous transform of the hand marker with respect to the shoulder $T_{\text{hand \_shoulder}}$ was used as a basis for the mapping protocol. Upon receiving an enable signal from the user (co-contraction of the forearm muscles), a snapshot of the current marker pose $T_{\text{hand \_ref \_shoulder}}$ is taken, along with the current end-effector pose with respect to the robot base $T_{\text{tool \_ref \_base}}$. The goal end-effector translation $t_{\text{tool \_goal \_base}}$ of the robot is then defined relative to the positional components of snapshots taken at enable time:

$$t_{\text{tool \_goal \_base}} = t_{\text{tool \_ref \_base}} + t_{\text{hand \_ref \_hand}}$$

The goal rotation of the end-effector $R_{\text{tool \_goal \_base}}$, however, is defined to directly mirror the orientation of the hand with reference to the shoulder frame:

$$R_{\text{tool \_goal \_base}} = R_{\text{hand \_shoulder}}$$

This relative translation and absolute rotation mapping allows intuitive control of the robot end-effector as the orientation of the reference frames feels natural from the user’s perspective. The resulting pose is filtered to reduce noise and exclude outliers before being passed to the inverse kinematics solver and robot controller. As an added safety measure, live robot tracking is immediately disabled on marker occlusion and can be re-activated by the user once markers are visible again.

IV. RESULTS AND DISCUSSION

The performance of the proposed framework was validated through three sets of experiments. The first set focused on evaluating the ability of the proposed learning algorithm to discriminate between different grasps during offline training, as well as online testing on the robot hand. The second part focused on evaluating the ArUco marker based pose tracking and mapping system for guiding the robot arm. The last part validated the complete, integrated telemanipulation system in a real-time pick and place task. All the experiments were recorded and the compiled video is available in HD quality at the following URL:

www.newdexterity.org/emgtelemanipulation

This study has received the approval of the University of Auckland Human Participants Ethics Committee (UAHPEC) with the reference number #019043. Prior to the study, the participating subjects provided written and informed consent to the experimental procedures.

A. EMG Based Robot Hand Control

The first experiment was aimed at evaluating the ability of the EMG based interface to discriminate between the following gestures: pinch grasp, tripod grasp, power grasp, muscle co-contraction, and rest state. The performance of three different classifiers (LDA, RF and SVM) was evaluated with these experiments. The final classification model was selected by considering the trade-off between classification accuracy and the time required to make the prediction. Tables I and II present the average time required to classify an individual sample and the resulting prediction accuracy of the three methods. The LDA classifier was found to be the fastest in terms of average prediction times, but achieved the lowest accuracy scores. The RF classifier had the highest prediction accuracies, while still offering prediction rates feasible for real-time operation. For this reason, the RF classifier was selected for the final framework implementation. Fig. 6 presents the confusion matrix over the 5-fold cross validation method for the RF based classifier for Subject 1. The trained classifier was tested in a real-time setting, where the robot hand was configured to mimic the user’s intended grasp type. The real-time experiment recordings are presented in the accompanying video.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Execution Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA</td>
<td>0.004</td>
</tr>
<tr>
<td>RF</td>
<td>0.021</td>
</tr>
<tr>
<td>SVM</td>
<td>0.100</td>
</tr>
</tbody>
</table>
### TABLE II
RESULTS OF CLASSIFICATION ACCURACY (A) AND STANDARD DEVIATION (SD) OBTAINED FOR THREE DIFFERENT CLASSIFIERS.

<table>
<thead>
<tr>
<th>Learning Model</th>
<th>Subject 1</th>
<th>Subject 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA</td>
<td>A (%)</td>
<td>93.57</td>
</tr>
<tr>
<td></td>
<td>SD (%)</td>
<td>3.27</td>
</tr>
<tr>
<td>RF</td>
<td>A (%)</td>
<td>97.90</td>
</tr>
<tr>
<td></td>
<td>SD (%)</td>
<td>1.01</td>
</tr>
<tr>
<td>SVM</td>
<td>A (%)</td>
<td>94.79</td>
</tr>
<tr>
<td></td>
<td>SD (%)</td>
<td>1.65</td>
</tr>
</tbody>
</table>

Fig. 6. Confusion matrix for the RF classifier. The x-axis represents the ground truth and the y-axis the classifier’s predictions. The diagonal represents classification accuracies for each arm hand posture class, where CC refers to co-contraction.

#### B. ArUco Marker Based Pose Tracking

The pose tracking and mapping component of the system was validated in three experimental settings.

1) **Head Motion Compensation**: This experiment tested the effects of user head motion on the robot end-effector pose. As the head moves, the camera view shifts and with it the estimated marker poses. This is compensated by referencing the hand marker to the shoulder and it is evident in the video that the target end-effector pose offsets are negligible even during large head tilts.

2) **One-Dimensional Motion Tracking**: In this experiment, the user moved their hand periodically in a single dimension to determine the teleoperation system delay. The motion consisted of three slow and three fast repetitions. The recorded goal and actual end-effector pose during the motion are presented in Fig. 7A. The robot arm tracking delay was measured to be approximately 0.2 s, which allows stable control as long as the user does not perform movement exceeding the critical frequency.

3) **Two-Dimensional Motion Tracking**: In this experiment, the user performed a circular motion with the aim of assessing the tracking accuracy and jitter. The recorded goal and actual end-effector pose during the motion are presented in Fig. 7B. It is evident that the robot motion follows its target quite closely and is not disrupted by the noise that is present in the goal due to marker pose estimation errors.

#### C. System Integration

In the final experiment, the system components were integrated and the complete telemanipulation framework was tested in a pick-and-place task. The user was able to grasp the object from the workspace and drop it in a bin successfully and with low effort. The experiment is best observed in the accompanying video.

**V. Conclusion**

This paper presented an intuitive telemanipulation framework for controlling a dexterous robot arm hand system based on electromyography and fiducial marker based pose tracking. The separate system components were validated and the complete framework was successfully tested in a remote manipulation setting.

Regarding future work, several aspects of the framework can be expanded and improved. In the next iteration, the g.Tec bioamplifiers will be replaced with a wearable EMG sleeve developed in-house \[37\]. To enhance pose tracking, additional markers could be placed on the operator’s arm.
to increase the tracking robustness and accessible range of rotation. The information from the fiducial markers and myoelectric activations can be merged to obtain kinematic and dynamic information for the arm pose. The framework is also planned to be expanded for bimanual arm hand system control, granting the operator better manipulation and re-grasping capabilities. The interface can be further enhanced by providing the head camera view angle information and grasp success feedback to the user. Finally, the framework ease of use should be assessed through trials on multiple subjects and compared to related teleoperation approaches.

REFERENCES


