Predicting Grasps with a Wearable Inertial and EMG Sensing Unit for Low-Power Detection of In-Hand Objects

Marian Theiss
University of Freiburg
Freiburg, Germany
theiss@ese.uni-freiburg.de

Philipp M. Scholl
University of Freiburg
Freiburg, Germany
pscholl@ese.uni-freiburg.de

Kristof Van Laerhoven
University of Freiburg
Freiburg, Germany
kristof@ese.uni-freiburg.de

ABSTRACT

Detecting the task at hand can often be improved when it is also known what object the user is holding. Several sensing modalities have been suggested to identify handheld objects, from wrist-worn RFID readers to cameras. A critical obstacle to using such sensors, however, is that they tend to be too power hungry for continuous usage. This paper proposes a system that detects grasping using first inertial sensors and then Electromyography (EMG) on the forearm, to then selectively activate the object identification sensors. This three-tiered approach would therefore only attempt to identify in-hand objects once it is known a grasping has occurred. Our experiments show that high recall can be obtained for grasp detection, 95% on average across participants, with the grasping of lighter and smaller objects clearly being more difficult.

CCS Concepts

•**Human-centered computing → Gestural input;**

Keywords

wrist-worn object detection; wearable EMG, inertial sensing

1. INTRODUCTION

In wearable gesture- and activity recognition, researchers aim at designing efficient and accurate systems that are able to detect what activities a person is performing automatically, by applying pattern detection and machine learning methods on the data of body-worn sensing units. These wearable units often are centered around inertial sensors that measure pose and motion; Such sensors have become not only popular because of their decreasing size and cost, they have also become very power efficient, allowing recordings of days to weeks on miniature batteries. This has led to numerous research platforms and systems being designed in the past decade, mostly to perform activity recognition in unconstrained settings, and often using just a single, miniature wrist-worn device. One obstacle that remains for wearable activity recognition with inertial sensors, however, is that for many physical gestures and motions, inertial sensors do not always exactly capture the essence of an activity.

AH 2016, February 25-27, 2016, Geneva, Switzerland

DOI: http://dx.doi.org/10.1145/2875194.2875207

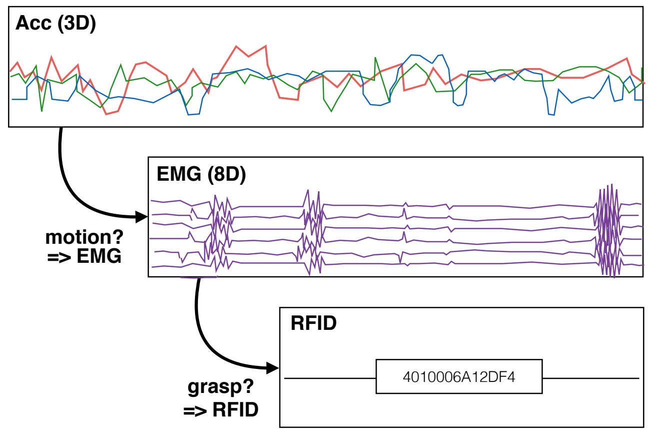

Figure 1: This paper presents a three-tiered system, in which grasps are detected through inertial and EMG measurements, after which the system activates the more energy-demanding unit for identifying what object or tool was grasped.

An object or tool being used by the user is often a valuable piece of information, in addition to what gestures are made (see e.g., [20]). By knowing what is being interacted with, i.e., what is being held in the user's hand, the estimation of user activity has been known to be improved considerably [14]. Such objects detection systems have been shown to work well in principle, though one major hurdle remains: The modalities available for detecting in-hand objects, such as wrist-worn Radio Frequency Identification (RFID) or cameras, tend to require a considerable amount of power for realistic deployments. This paper focuses on a system design (Figure 1) that can efficiently detect the type of object that is held in the hand, by using a wearable RFID scanner in combination with an inertial measurement unit (IMU) and eight EMG units placed on the lower arm. Since the searching of nearby RFID tags is the most energy-demanding sensing tasks, we propose to minimize this action only to instances when the user is about to grab a new object, thus leaving the RFID-scanning mostly turned off. Grasping of an object is only tested for when sufficient evidence is provided from the more power-efficient IMU.

This paper is structured as follows: After situating our research within related research in grasp and object detection approaches, we present our system's main concepts and system implementation. The following section then will focus on the selection and data analysis of the two specific sensor modalities, EMG and inertial data, after which our experiments and their results will be presented. The final conclusions section will then present our main findings and suggest several future directions for this work.

2. RELATED WORK

Research in HCI and wearable systems has proposed a wide diversity of approaches that are capable of automatically detecting objects in the user's hands. The motivation for systems that detect in-hand objects can range from improving basic interaction by using tangible and familiar tools [3, 19], to tracking of where users might have left objects from where they last grasped them [17], to obtaining and tracking the user's daily activities through any activity-dependent held objects [9, 13, 14, 15]. Earlier work of wearable object detection proposed the use of RFID tags combined with a tag reader in a wearable setting to detect basic interactions and activities [3, 4, 5, 9, 13, 14, 16, 19]. More recently, wearable research has suggested alternative, but equally power-demanding sensing modalities such as wrist-worn camera systems as a basis for performing in-hand object detection from wrist-worn devices (as in [1, 11, 12, 17]).

One of the challenges in designing wrist-worn RFID readers is that tags need to be detected from a relatively large range (between 10 and 25 cm are reported in the aforementioned research), requiring either a large and well-calibrated antenna (as for instance detailed in [3]) or amplification. The latter implies that either more battery power is needed on the receiver's end, i.e., on the wrist, or that an additional battery is provided at every tagged object. Want remarks in [18] that passive RFID, where tags are kept cheap, small, and robust since they lack a battery, is by far the more interesting class.

For camera-based object detection from the wrist, two obstacles present themselves from an energy perspective: First, the camera needs to perform constant capturing of images, which even for miniature and low-resolution cameras requires about 80-100 mA of current consumption[1].

Second, the images need to be analyzed for potential objects present at different scales, orientations, and being possibly partially occluded by the user's hand. A major benefit of this modality is that objects do not have to be tagged beforehand and that systems as presented in [11], [12] and [17] could be used with a larger diversity of objects. The combination of these two requirements, however, means that the power requirements, along with the components' cost and size, are very large.

A third alternative modality for sensing in-hand objects that has been suggested recently is capacitive sensing [7, 8] from wrist-worn electrodes; Although this approach characterizes object types only *roughly* through capacitive proximity sensing, the authors have shown that recognition of interaction and activities is improved when adding this information. The lack of exact object identification in this work is countered by this approach's considerably more power-efficient operation, with a reported 1 mA draw at 3.3V.

Early work by Schmidt et al. [16] suggested a wearable RFID reader for detecting the objects in the environment as they are handled by the user. Their design suggested the implementation of the antennas as coils inside clothing, in particular gloves. They also show how the detected tags within objects can trigger applications implicitly, and suggest use cases in inventory managment and data warehousing. Work by Fishkin et al. [5] followed up on such work, showing how to integrate wearable RFID readers in gloves (iGlove) and suggesting wrist-worn, bracelet-like (iBracelet) systems as an alternative and less-obtrusive solution.

Philipose et al. [14] demonstrated the possibility of detecting activities of daily living with wearable object detection. They used a glove based RFID reader with a sampling rate of 2 Hz which lasted for 2 hours. They collected a series of tags, corresponding to an activity, and calculated the probability that one activity is executed. Besides modeling errors, they also identified sensor errors (missed tags) as a source of ambiguity. They achieved a recall of 73% with a precision of 88%. They also reported a recall of 33% for their worst recorded activity. As a reason for missed detections, they mentioned the absorption of the RFID signals by the environment. Patterson et al. [13] showed in a similar setup for fine-grained daily activity recognition, that incorporating the knowledge of aggregate objects leads not only to good accuracy but also requires less training data to learn the activties' models.

Berlin et al. [3] designed an open-source RFID reader with an oval, wrist-worn antenna. The project addresses the questions of increasing the range of an antenna, obtaining the optimal sampling frequency for the reader, and how 3D acceleration patterns can be approximated efficiently for detecting gestures of interest. Because of the relatively large distance of the wrist to the tagged objects in the hand, the RFID antenna was enlarged and tuned. The antenna and RFID reader unit was described with 60 mA while reading, 15 mA in idle mode, and 60 µA in sleep state. One reading needed between 20 and 68.4 ms, depending on tag type and whether the reading is successful. With one reading per second, it was possible to decrease the average current consumption to approximately 18.23 mA. In the tests, the hit rate started at 100% with 16 Hz (full duty cycle) and decreased to 65% with 1 Hz.

An approach to use wearable in-hand object detection to interact with the environment was shown by Wolf et al. [19], who introduced an interaction device with a wearable object pick-up detection system. The work sought to switch devices from standby to active mode when taking into hand instead of a manual activation technique. The system uses a combination of a RFID ring with an embedded gyroscope ("PickRing") and the gyroscope data of the devices. Gyroscope data is sent via Bluetooth to all coupled devices and compared with their gyroscope data. The entirely wearable part of the system, with microcontroller and Bluetooth active, has a reported runtime of 15 hours on a 9V battery.

Common to all these research efforts is that object detection from the wrist is shown to have a lot of advantages and would open up new interaction possibilities or improved recognition and tracking systems. Power efficiency, however, is without exception identified as a main obstacle to achieve such systems in real-world applications. In the following sec-

[1]This estimate was taken from a range of commercially available 2M JPEG Color Camera modules with TTL level interface.

tions, we will describe our proposed approach, which adds two sensing modalities to make such object detection systems more power-efficient by turning the in-hand object detection on, only when actual grasping gestures are spotted.

3. SYSTEM CONCEPTS

Since the runtime of wearable activity sensors is dictated mostly by their battery, achieving a minimal power consumption is a primary goal. Furthermore, obtaining the sensor readings requires a significant portion of current systems that are able of detecting in-hand objects. To this end, we argue for adding a mechanism to leave the object detection unit turned off when no objects are expected to be in the user's hand, to decrease the power consumption of the entire system. As a result, such systems could be made reduced in size and weight (not unimportant, since they have to be worn at the wrist) and longer-lasting on a single battery charge.

3.1 Method Overview

The detection of grasping is thus added to the system, to enable to flexibly activate object detection as soon as a grasp is detected (as depicted in Figure 1). It is important to note here that adding this additional system only makes sense if this in turn does not require a similar amount of power. We propose to employ in our system a combination of EMG sensors and a low-power IMU unit, which combined need much less power than either RFID or camera-based units. We furthermore use the same technique in the grasp detection part of the system: The only sensor which is consistently active is the accelerometer for motion detection. As soon as sufficient motion is detected, the EMG sensors are switched on to guarantee a sufficiently accurate detection of grasps.

As in most classification problems, our system's detection performance cannot be expected to be flawless. It is more critical that all grasps are recognized, while occasional false positives (i.e., when grasps are detected when the user was not grasping an object) are less problematic. From a modeling perspective, the aim is thus to maximize the recall performance of the detection system, while allowing for less optimal precision figures. This would result in a system that sporadically has the RFID tag reading unit activated, at every occasion once a grasp gesture is detected. The system loses some energy whenever other gestures are misclassified as grasps, but this we argue is much better than missing grasps (and therefore objects) altogether. Whether we are able to construct our detection system in this way, is what will be focused on in the evaluation section. First, though, we will describe the used hardware in the next section.

3.2 Implementation Details

Our grasp detection system's prototype is based on EMG and IMU data. As sensor we used a commercial system, the Myo from Thalmic Labs[2]. The IMU sensor contains a 3D accelerometer, a 3D magnetometer, and a 3D gyroscope, all sampled at 50 Hz. EMG data is collected via 8 units at 200 Hz. The EMG units are worn on the upper part of the forearm (see Figure 2), a few centimeters away from the elbow. All EMG electrodes are equidistantly distributed around the forearm. The Myo has a built-in Bluetooth Low-Energy (BLE) unit that can be accessed as a serial port

[2]https://developer.thalmic.com, [last accessed Dec. 2015]

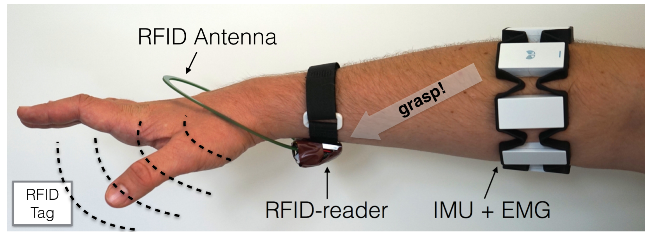

Figure 2: Our prototype uses a commercial EMG and IMU sensor (Myo from Thalmic Labs, depicted right), which is through BLE connected to a wrist-worn RFID reader (left). Only once a grasping gesture is detected from IMU and EMG data, our system instructs to activate the RFID reader and read any nearby tags, via BLE.

from another host unit. In our case, we used the design from [3] to obtain an RFID unit with a wide reading range, with an integrated RFduino[3] processing unit that includes a BLE tranceiver as well. This way, both modules form one system, with the main point of control at the RFID reader side (due to the fact that the Myo's microcontroller cannot be altered), being able to function for about 9 hours on full battery charges. To avoid the constant sending of sensor data between the two modules, and thus obtain a much more energy-efficient configuration, one could move this control to the Myo's side. This paper's focus is however in the grasp detection methods and the reliabilty thereof, for which this system's runtime is more than adequate enough.

For the experiments, the logging of the raw sensor data has been implemented via a C++ program on a laptop nearby connected via BLE to the system. Matlab from MathWorks has been used for postprocessing the data, extracting the features, and visualization. A real time grasp detection method has also been developed in Matlab. Based on the specific example of the hardware prototype described above, the next section will enumerate the typical energy consumption figures for each of the individual components in the system, to motivate why our proposed three-tiered approach can be expected to lead to considerable power savings.

3.3 Energy Requirements for Components

With the prototype system in place, using state-of-the-art components, we can now investigate the current consumption of each of the components as an illustration of the energy saving opportunities that lie in the appropriately switching off of sensing modules (RFID and EMG) when they are not needed. For typical energy consumption, the following estimations are made, based on available datasheets of the main components and with an operating voltage at 3V:

- Accelerometer (InvenSense MPU9150): ...... $<$0.2 mA
- Gyroscope (InvenSense MPU9150): .... ~3 mA
- 8 channel EMG (based on ST 78589 VA1814): .. ~8 mA
- Processing (Freescale K21 Cortex M4 MCU): . ~7.7 mA
- RFID (SkyeTek SkyeModule M1-Mini): .... ~60 mA

[3]http://www.rfduino.com/ [last accessed Dec. 2015]

The estimates show that for each additional tier, the sensing modules draw one order of magnitude more current than those in the one before: Accelerometers typically draw far less than 1 mA, the EMG sensors are estimated to require around 8 mA, while an active RFID reader uses 60 mA.

The first layer of our grasp detection model is based on acceleration data alone and observes the patterns present in the data from the inertial sensor to test whether they would support the presence of grasps. The underlying idea here is that grasping is unlikely to take place when the user is not moving his hand (such as during resting or sleeping) or when an activity can be recognized that inhibits grasping (such as while walking or running). Since the gyroscope data has in previous experiments led to similarly potent features as the acceleration data, the 3 mA can be neglected by using solely the accelerometer for detecting sedentary behavior and specific activities that inhibit grasping. By thus neglecting the gyroscope's influence and assuming a normal resting and walking activity, the system would draw an overhead current of 12 mA.

Energy consumption of RFID readers is mostly depending on their duty cycle (reading all nearby tags) and possible detection range. With a full duty cycle over short ranges, a nominal current consumption can reach over 100 mA. In [3], the M1 with a wrist-worn antenna design was shown to consume 18.23 mA during a low 1 Hz duty cycle, although tags were much more reliably read at 16 Hz with a higher current consumption. Improvements could include using a more powerful RFID reader unit between the RFID antenna, leading to larger units that require more energy, however.

4. THE GRASP DETECTION MODEL

For the detection of the grasps, a three-layered model has been implemented to decrease the level of abstraction from a detection in the whole system to a detection in smaller subsystems. In the first layer, we perform a basic type of activity recognition, where a person's activity is estimated for a period larger than one second. Examples for such situations are resting, walking or working. In the second layer, typical features for the actual task of grasping are extracted. These features are calculated over a time period of a few hundred milliseconds to achieve a high reactivity of the system. Because of a wide variation of grasps and movements, similar to grasps, single features by themselves tend to be rather weak. In the third layer, the chronological sequence of detected features is matched and classified. This layer ensures a good combination of the detected features which supports the grasp estimate, and therefore keeps the precision high. The high precision allows to collect a lot of weak features in the second layer which enables the system to detect a wide variation of grasps to achieve a high recall in real world applications. To highlight where the features, especially those from the EMG data, occur during the grasping sequence, we will in the next section list the main steps during the grasping of an object.

4.1 Steps for a Grasping Sequence

In the third layer of the model, the grasping action can be divided into different steps, with each of the steps invoking different features. Typically, however, there is no clear border between these steps that would make it easier to flag the detection of a feature for a period of time. For feature extraction, the steps "Moving hand to object", "Closing the hand around object", "Holding object" and "Releasing object" were used. The collection of features for a step is typically detected within a short time interval, but different steps can be spread in time. For this situation, it can be expected that features before a grasp are further away from the actual grasp than features for the holding step. The steps have been depicted in Figure 3: The pictures demonstrate a typical sequence for a grasp. The first picture (Figure 3.1) illustrates a resting posture just before (and typically also present after) the grasp. Data associated to this step should not be detected as part of a grasp, and is thus part of the null class.

Moving hand to object is shown in Figure 3.2. This step can use motion features for changing the position of the lower arm, a directed movement of the arm and hand towards the object and, very typically, also a rotation of the hand. The EMG features are typically representing the opening of the hand and a slight pulling back of the back of the hand.

Closing hand around object is shown in Figure 3.3. This often tends to contain strong accelerometer signals going in the opposite direction of the previous movement by the slowing down of the hand or its colliding with the object. Specific muscle activation for a flexion can also be detected.

Holding object is illustrated in Figure 3.4. This step of the grasping action can be detected by continuous strong signals coming from muscular activity used for closing the hands. Light-weight and fragile objects can be problematic to detect in this step since a strong grasp is not necessary or impossible.

Releasing object is shown in Figure 3.5. Typically, similar features to the "Moving hand to object" step are invoked during this step. Since this occurs after the grasp has taken place, it has no direct use to activate the RFID reader. However, it does contain valuable information about the ending of the grasp (and therefore stopping the object detection).

For extracting features, the eight EMG signals, the three acceleration channels and three gyroscope channels were analyzed in the time domain, as well as in low frequency energy and entropy and in (spectral) frequency bands. The acceleration data was both represented with and without the gravity factor (to get rid of the gravity factor a high pass filter is used).

4.2 Feature Selection

Since the hardware prototype contains a full IMU, we initially evaluated the gyroscope's output, as well as using the IMU's pre-computed Euler angles. Surprisingly, the resulting features from these modalities were found to be less distinctive than those from the basic 3D acceleration modality. Inertial measurement units have the advantage that they are more common, cheaper to integrate into designs, and smaller and thus more comfortable to wear. However, in case of grasping, IMU signals alone were found to be not specific enough to accurately detect grasping gestures in our first studies.

For the EMG features, we used methods similar to those used in related work such as [10] and [2]. In case of EMG features, it is helpful to get an idea about the muscle positions and functions. Roughly spoken the extensors of the wrist and fingers can be used to get the step before a grasp, while the flexors are a better detection for the grasp itself. Flexors are typically more on the anterior and extensors are

Figure 3: Illustrating the sequence of steps during a complete grasp. The figures illustrate the modelled steps during the grasp, as also indicated in the text and in Figure 4: (1) Start/stop posture, (2) Moving hand to object, (3) Closing hand around object, (4) Holding object, and (5) Releasing the object.

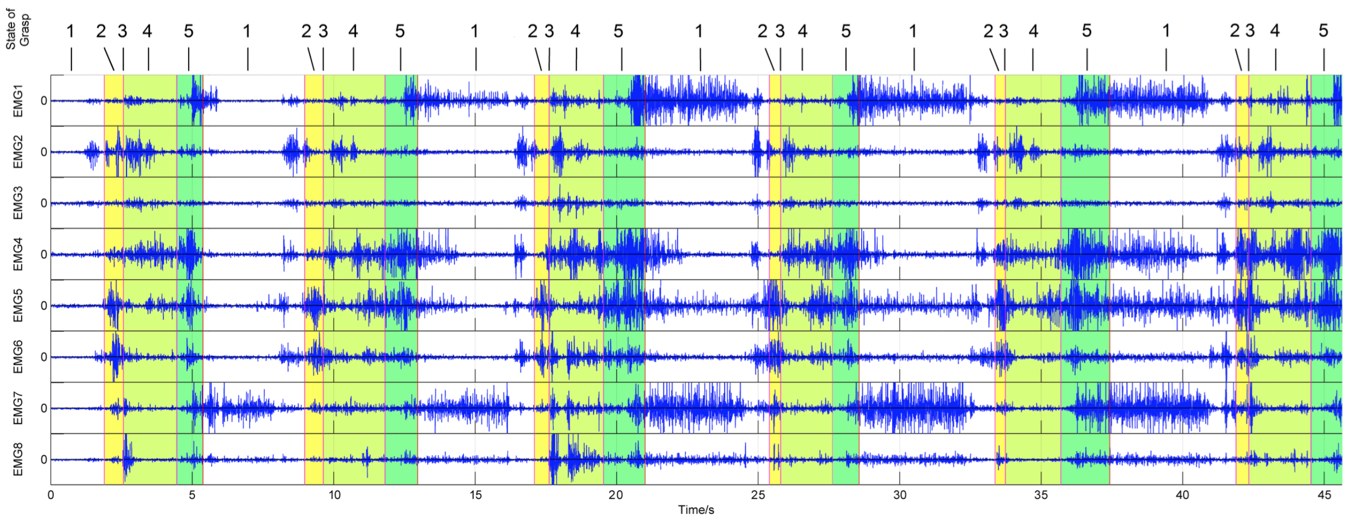

Figure 4: Activation signals of the eight EMG sensor units over time in a series of seven grasps of different objects, from the same participant. Grasping steps are numbered according to the steps in Figure 3. Colored bars indicate a step change, where step 3 is a relatively short period of time. The EMG sensor units have been worn equidistantly on the right forearm, near the elbow, with a clockwise numbering starting on the anterior side.

more on the posterior side of the forearm. Each of the features should be checked statistically across multiple persons, because some muscles only show an activation under a lot of stress. It is important to note that measurements on the skin can also lowpass-filter the signals and decrease the signal strength. As such, EMG features should be rather checked for an activation than for static values, since the EMG signals not only differ by person but also for the duration of wearing the sensors.

Figure 4 shows a series of seven typical grasps. The numbering on top of the figure matches the numbers in Figure 3. The EMG sensors are worn equidistantly around the right forearm, around 4-5 centimeters away from the elbow. The numbering of the sensors is made clockwise: Sensors 8, 1 and 2 are worn on the anterior side and sensors 4 till 6 are worn on the posterior side of the arm. Steps 1, 2, 4 and 5 are shown between the bars, step 3 is a very short time period marked with the second magenta bar of each grasp. Step 1 is part of the null class and is not used when building a detector, however the features should be distinguishable from it. Step 2 is the arm movement to the object with the intention to grasp: Usually, sensors 5 and 6 show a voltage increase because of the extension of the wrist and fingers. Step 3 marks the closing of the hand around the object: Mostly, short spikes in the voltages of the flexor sensors 1 and 8 can be seen. Step 4 marks the holding of the object: Depending on the strength of the grasp, all sensors, but mainly sensor 4, show an increased voltage. Step 5 marks the letting go of the object which can, but must not, start with a flexion of the wrist, shown by a spike in the corresponding sensors and then an extension with an increased voltage in sensor 4, 5 and 6.

The IMU features can be derived by the typical arm movements. Usually a directed movement towards the object, a very common twist of the arm and a change of the location of the hand (seen in the gravity factor of the acceleration) can be detected. Features were here limited to a window of 0.2 seconds and consisted of the basic statistics (minimum, maximum, difference thereof, mean and variance) in the time domain. The required calculation frequency of the features should not be too fast to save processing power but also shouldn't be too slow in case of short grasps. Assuming a minimum grasping length of 650 ms and a maximal searching time for tags of 70 ms, a mode switching delay of the reader of 100 ms and a maximum of 2 calculation steps after the grasp, a delay between the calculations of 220 ms is appropriate. For feature selection and classification, we opted for using decision stumps as a basic classifier, with finally detecting the grasping by a boosting approach. As an alternative, PCA has been tested to decrease the feature space, but it was found that it needs roughly four till six components to reach a certainty of above 95% for classifying the step of the grasping gesture. We finally also evaluated several other classification methods available through GRT [6] in leave-one-participant-out cross-validation, though these did not deliver better results.

5. EXPERIMENT

In order to evaluate how well a system such as ours can generally detect grasps, especially as these grasps are underway, we will present in this section a feasibility study, using our hardware prototype in a set of controlled experiments.

5.1 Experiment Setup

For the evaluation, a set of scripted tasks were performed by twelve participants. They were told before the experiment how to wear the hardware prototype and to try to use their right hand for all tasks. In order to achieve an as natural as possible behavior from the participants, they were not told that the experiment's goal was to monitor grasps specifically, and tasks were instead about fetching objects from a variety of places, at a variety of heights, with interleaved activities such as walking or drinking from a bottle. No instructions on how to move the arm or how to perform grasps was made, so the participants were free to use any grasping sequence.

Before each experiment, care was taken that the orientation of the IMU unit and the placement of the EMG electrodes were near-identical for each participant. The experiments did not contain a calibration phase, the sensors were used straight away. The task included RFID-tagged objects as similar as possible to the box test reported in [3], including a cardboard box including a hammer, a screwdriver, a craft knife, a filled 0.5-litre plastic water bottle, an empty 1-litre plastic bottle and a USB stick. These items are also depicted in Figure 5.

The first two tasks involved the grasping interleaved between walking around in the room. For this, the participants were asked to walk to a water bottle, grab the bottle, bring it back, drink a bit and put it on the table. Afterwards they were told to bring a cardboard box to the table, filled with tools needed for the next tasks. After this, the participants were asked to perform several basic tasks, involving getting several objects out of the box and putting them back after use. A nail was for instance hammered into a wooden board, a screw was screwed, some cuts were made in a paper sheet, other objects like the empty bottle or the USB stick were moved around between box and table. Afterwards the box was put on its side so that the opening pointed towards the participant, after which several more tasks followed.

As a baseline method for comparing against, we chose to use a basic classifier from the Gesture Recognition Toolkit [6] using the Random Forest Algorithm to train for individual steps on a range of features (mean, range, variance, median, zero crossings, root mean square) that were z-normalized.

5.2 Experiment Results

For evaluating the proposed system, the log files of the experiments have been postprocessed with the same configurations as one would use for real-time analysis. The timestamps of the recognition have been synchronized with ground truth annotations from video data taken during the experiments. For this ground truth, we used the time when

Figure 5: Selected items used in the experiment were tagged with RFID tags and were chosen to be similar with the objects reported in [3].

the hand or fingers start touching the object with enough force so that the object would not fall off when moving. For an accepted true positive instance, the found timestamp had to be between 50 ms before and 350 ms after the ground truth mark.

Per participant, a total of 18 grasps should have occurred. Because of a relatively loose usage of the objects and due to extra gestures between the grasps, this number was often slightly higher. For instance, some participants were putting down the hammer first for grasping a nail. Grasping other objects, like the nails, has also been detected but since it was sometimes hard to distinguish whether a touch should be counted as a grasp, only moving of the mentioned objects was counted. Other gestures for which it was hard to decide whether they could be counted as grasps, were for example the laying down of the hand on the leg.

Since the number of grasps differed per participant and per object, we received three different recall rates. A recall of 95.20% was achieved when counting every grasp, 95.50% was achieved when taking the mean recall over the participants for every grasp and 95.54% when taking the mean of the objects' recall results for all participants. The single results can be seen in Figure 6. Five of twelve participants received a recall of 100%. The two worst results were close above 85%. Recall results for the objects can be seen in Figure 7. The cardboard box and the filled water bottle received the best results with 100%. The more light-weight or easier to handle objects (USB stick, screwdriver and craft knife) received the worst results between 89.29% and 92%.

Results are primarily expressed as recall, as detecting every grasp is the most crucial for our system. A slightly lower precision (all true positives versus all positives) is not as critical, as it will 'just' cause the object detection system to be activated needlessly (as there is no object to detect). As long as this does not occur too frequently, the impact on the entire system's energy consumption is minimal. Most false

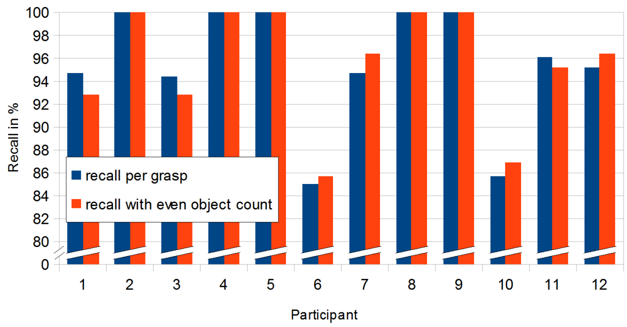

Figure 6: Recall for each of 12 participants with all used objects, either across all grasps (red) or distributed evenly over the objects (blue).

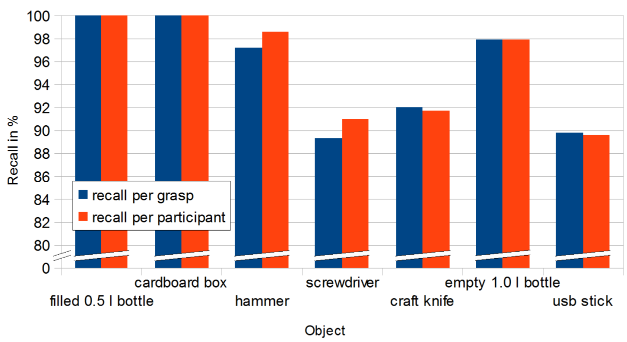

Figure 7: Recall results for grasps of the objects across all participants, either across all grasps (red) or distributed over participants (blue).

positives were furthermore found to occur around the grasp, when performing strong movements or when strengthening a grasp.

5.3 Experiment Discussion

Using our proposed set of IMU and EMG features generally performed better than the baseline classifier (which resulted in an across-objects recall range of 68-84% and precision range of 68-78%, for the 12 study participants).

The false positives can be categorized into three main classes. Right before and after the actual grasp, often some detections can be found. This number could be reduced by disabling the detection for some time after finding an object by the object detector. Strong movements of the arm can also be detected as grasps if the walk detection does not filter them out. Strengthening of a grasp is also often detected. This can also get filtered out by a disabling after object detection or by checking for an end of the grasp before enabling a new detection.

A system based on acceleration (or gyroscope data) alone is hard to implement since the found accelerometer (and gyroscope) features are hard to distinguish from many other movements. Moreover, the change in the time signals of a grasp movement is much weaker than most other movements. The most problematic issue with using IMU data only is that the step "Holding object" is hard to model. Therefore, a high recall is only possible with a very low precision, which would result in object detection that is nearly always on.

Some grasps differ strongly from the typical grasps in the resulting sensor values and therefore should use different features. To overcome this problem, more universal features can be used which results in a higher false positive rate. An alternative can be a second detection routine with different features. Grasps of light-weight and fragile objects are harder to be found. Persons are more careful when moving towards these objects and only weak acceleration changes are found. Since persons needs less strength or be more careful when touching these objects, the EMG data do not have such strong differences in the signal. Lastly, when tightening a grasp slowly, the short-time EMG signals do not differentiate much. To detect grasping this type of objects, the detector has to react on weaker changes, which would increase the false positive rate and would thus lead to a less efficient system.

Even if we tuned the parameters generally with an acceptable result, the functionality of the system is user-dependent. It is entirely possible that people might prefer grasps that are less typical and harder to detect. The EMG signals differ per person and as the hardware is worn. The IMU parameters worked well in the tests, but it is possible that they have to be altered for some persons. We also found during initial experiments that participants act unnaturally when getting immediate feedback. It could be possible that this reaction can also be observed when using it in real life applications. Achieving a fixed sensor orientation is straightforward, but placing the EMG sensors on identical positions on the forearm across all participants is hard to achieve. A more accurate position would result in better responses, so designing the sensors to fit uniformly (e.g., from a fixed distance from the elbow), or having feedback for EMG signal quality while fitting would likely increase grasp detection as well.

6. CONCLUSIONS

A system that is able to identify objects in the user's hand has been suggested for interaction, tracking user objects, and improving the recognition of user activities. A major obstacle remains that such systems are far from energy-efficient; The cheapest solution that will deliver immediate object identification, using wrist-worn RFID readers, requires still too much battery power for realistic or non-obtrusive deployment. In this paper we have focused on a solution that is based on the addition of two modalities that are an order of magnitude cheaper in terms of power consumption: By first detecting whether grasping is taking place by means of an IMU and an array of EMG sensors, the RFID reading is only switched on when the user makes a grasping gesture.

We developed a grasp detection system based on efficient analysis of EMG and IMU data, which is connected to an RFID reader unit via BLE. In this paper, we focused on the ability of a grasp detection system to correctly identify grasps with a priority on spotting every grasp, allowing also for the occasional false positive (as this turns the RFID reader on for a short time although no grasp occurred).

In a set of experiments, we asked 12 participants to wear our system and perform a series of tasks, including several grasping tasks from a variety of positions, with interleaved other activities, and with a variety of objects to grasp. The overall recall (grasps that were detected by our system) was above 95%, with 100% for five out of twelve participants and also for the heavier two out of seven objects. We found that false positives (motions that were detected as a grasp but were not) are limited and can be categorized in three classes: Those that occur very close to the grasp, those that are associated with single strong arm gestures, and those that occur when strengthening a grasp. We identified two hurdles that remain before such a system can be adopted: First, grasping light-weight objects tends to be less reliable to detect. Second, the effectivity of grasp detection with EMG fluctuates strongly across persons due to skin type, wear, and positioning.

A real time version of the prototype has also been implemented, which gives vibration and visual feedback for detected grasps. We are currently implementing the presented features and detection routines on an embedded hardware prototype similar to the one presented here, and are investigating possible ways of integrating the whole system in a single wrist-worn unit, so that the BLE link present in the current prototype can be avoided. This would enable future research to include both more realistic and more prolonged user experiments.

7. ACKNOWLEDGMENTS

We thank all the volunteers of the user study for their participation in the experiments, their feedback and patience.

8. REFERENCES

[1] F. Ahmad and P. Musilek. A keystroke and pointer control input interface for wearable computers. In *Fourth Annual IEEE Intl. Conf. on Pervasive Computing and Communications*, pages 2–11, 2006.

[2] O. Amft, H. Junker, P. Lukowicz, G. Troester, and C. Schuster. Sensing muscle activities with body-worn sensors. In *BSN 2006: Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks*. IEEE press, 2006.

[3] E. Berlin, J. Liu, K. Van Laerhoven, and B. Schiele. Coming to grips with the objects we grasp: Detecting interactions with efficient wrist-worn sensors. In *International Conference on Tangible and Embedded Interaction (TEI 2010)*, pages 57–64, Cambridge MA, USA, 01/2010 2010. ACM Press, ACM Press.

[4] A. Feldman, E. M. Tapia, S. Sadi, P. Maes, and C. Schmandt. Reachmedia: On-the-move interaction with everyday objects. In *Proceedings of the Ninth IEEE International Symposium on Wearable Computers*, ISWC '05, pages 52–59, Washington, DC, USA, 2005. IEEE Computer Society.

[5] K. P. Fishkin, M. Philipose, and A. D. Rea. Hands-on RFID: wireless wearables for detecting use of objects. In *Ninth IEEE International Symposium on Wearable Computers (ISWC 2005), 18-21 October 2005, Osaka, Japan*, pages 38–43. IEEE Computer Society, 2005.

[6] N. Gillian and J. A. Paradiso. The gesture recognition toolkit. *The Journal of Machine Learning Research*, 15(1):3483–3487, 2014.

[7] T. Grosse-Puppendahl. *Capacitive Sensing and Communication for Ubiquitous Interaction and Env. Perception*. PhD thesis, TU Darmstadt, May 2015.

[8] T. Grosse-Puppendahl, E. Berlin, and M. Borazio. Enhancing accelerometer-based activity recognition with capacitive proximity sensing. In *Ambient Intelligence*, volume 7683 of *LNCS*, pages 17–32. Springer, 2012.

[9] A. Hein and T. Kirste. A hybrid approach for recognizing adls and care activities using inertial sensors and rfid. In C. Stephanidis, editor, *Intelligent and Ubiquitous Interaction Environments, 5th International Conference, UAHCI 2009*, Lecture Notes in Computer Science, pages 178–188, San Diego, CA, USA, 2009. Springer.

[10] A. Latwesen and P. E. Patterson. Identification of lower arm motions using the emg signals of shoulder muscles. *Medical Engineering and Physics*, 16(2), 1994.

[11] T. Maekawa, Y. Kishino, Y. Yanagisawa, and Y. Sakurai. Wristsense: Wrist-worn sensor device with camera for daily activity recognition. In *Pervasive Computing and Communications Workshops, 2012 IEEE Intl. Conf. on*, pages 510–512, March 2012.

[12] T. Maekawa, Y. Yanagisawa, Y. Kishino, K. Ishiguro, K. Kamei, Y. Sakurai, and T. Okadome. Object-based activity recognition with heterogeneous sensors on wrist. In *Pervasive Computing*, volume 6030 of *Lecture Notes in Computer Science*, pages 246–264. Springer Berlin, 2010.

[13] D. J. Patterson, D. Fox, H. A. Kautz, and M. Philipose. Fine-grained activity recognition by aggregating abstract object usage. In *Ninth IEEE International Symposium on Wearable Computers (ISWC 2005), 18-21 October 2005, Osaka, Japan*, pages 44–51. IEEE Computer Society, 2005.

[14] M. Philipose, K. P. Fishkin, M. Perkowitz, D. J. Patterson, D. Fox, H. Kautz, and D. Hahnel. Inferring activities from interactions with objects. *Pervasive Computing*, 3(4):50âĂŞ57, 2004.

[15] J. Rekimoto. Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In *Proceedings of the 5th IEEE International Symposium on Wearable Computers*, ISWC '01, pages 21–, Washington, DC, USA, 2001. IEEE Computer Society.

[16] A. Schmidt, H. Gellersen, and C. Merz. Enabling implicit human computer interaction: A wearable rfid-tag reader. In *Fourth International Symposium on Wearable Computers (ISWC 2000), Atlanta, Georgia, USA, 18-21 October 2000, Proceedings.*, pages 193–194, 2000.

[17] F. Tavakolizadeh, J. Gu, and B. Saket. Traceband: Locating missing items by visual remembrance. In *Proceedings of the Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST'14 Adjunct, pages 109–110, New York, NY, USA, 2014. ACM.

[18] R. Want. An introduction to rfid technology. *IEEE Pervasive Computing*, 5(1):25–33, Jan. 2006.

[19] K. Wolf and J. Willaredt. Pickring: seamless interaction through pick-up detection. In *Proceedings of the 6th Augmented Human International Conference, AH 2015, Singapore, March 9-11, 2015*, pages 13–20, 2015.

[20] A. Zinnen, K. Van Laerhoven, and B. Schiele. Toward recognition of short and non-repetitive activities from wearable sensors. In *Proceedings of the 2007 European conference on Ambient intelligence*, pages 142–158. Springer-Verlag, 2007.