*O*Sense: Object-activity Identification Based on Gasping Posture and Motion

Thisum Buddhika[1], Haimo Zhang[2], Chamod Weerasinghe[2], Suranga Nanayakkara[2] and Roger Zimmermann[1]

[1]National University of Singapore, School of Computing, Singapore
{thisum, rogerz}@comp.nus.edu.sg
[2]Augmented Human Lab, Auckland Bioengineering Institute, The University of Auckland, New Zealand
{haimo, chamod, suranga}@ahlab.org

Figure 1: *O*Sense propose a new technique to identify object-activities by using motion data and posture data. This figure shows the object-activities used to evaluate *O*Sense.

ABSTRACT

Observing that, how we grasp objects is highly correlated with geometric shapes and interactions, we propose the use of hand postures and motions as an indirect source of inputs for object-activity recognition. This paradigm treats the human hand as an always-available sensor, and transforms all sensing problems to the data analysis for the "sensor hand". We envision this paradigm to be generalizable for all objects regardless of whether they are acoustically or electromagnetically active, and that it detects different motions while holding the same object. Our proof-of-concept setup consists of six IMU sensors mounted on the fingers and back of the hand. Our experiments show that when the posture is combined with the motion, the personalized object-activity detection accuracy increases from 80% to 87%.

AH2019, March 11–12, 2019, Reims, France

ACM ISBN 978-1-4503-6547-5/19/03…$15.00
https://doi.org/10.1145/3311823.3311841

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing systems and tools**; *Interactive systems and tools*;

KEYWORDS

Wearable Computing, Object Identification, Activity Recognition

ACM Reference Format:
Thisum Buddhika[1], Haimo Zhang[2], Chamod Weerasinghe[2], Suranga Nanayakkara[2] and Roger Zimmermann[1]. 2019. *O*Sense: Object-activity Identification Based on Gasping Posture and Motion. In *Augmented Human International Conference 2019 (AH2019), March 11–12, 2019, Reims, France.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3311823.3311841

1 INTRODUCTION

We extensively use our hands to grasp and interact with objects in everyday life. Clues about how an object should be used (e.g. the way to grasp the object) are typically hinted by the shape of the object and its context. Research on neuroscience [2] and biomechanics [15] of the human body shows based on the object shape and size, the hand is pre-shaped even before the hand is grasping an object. Further studies have shown the influence of object shape on hand posture, irrespective of handedness [5].

Motivated by this, we propose a new technique to identify object-activities by using motion data and posture data. To capture data, we developed a hand-worn prototype, *O*Sense, that has six Inertial Measurement Units (IMUs). To capture the activity information, we used 3-axis accelerometers of the IMUs. To determine the posture, we used sensor fusion method to calculate relative quaternions incorporating the accelerometer, gyro and the magnetometer. *O*Sense system was developed and evaluated with data from 12 users for 8 objects and 9 interactions including *No-Action*. We identify object-activities primarily using motion data and use the posture data to disambiguate between confusing motions with different objects. The main contribution of our work is an empirical study to show how the object identification can be done with motion and posture data.

2 RELATED WORK

Among various types of object recognition techniques, computer vision is the widely used method. Previous work has been done by wearing cameras on different parts of the body [3, 16]. Although many objects can be identified with high accuracy compared to other existing methods, line-of-sight, occlusion and lighting conditions are the major issues with these systems. Maekawa et. al. [11] used glove-based hand-worn magnetic sensors to detect handheld electrical devices by sensing the time-varying magnetic fields emitted while operating. EM-Sense [7] used electromagnetic noise sensing method to detect electro-mechanical objects while operating. ViBand [6] used a modified off-the-shelf smartwatch to sense the vibrations emitted by the devices when operating. The main limitation of these two approaches is they can only identify objects having specific characteristics while operating. On the other hand, tag-based methods such as RFID [4] can be used on any object, but have limited scalability, as every object needs to be tagged.

Object recognition vs. Grasping: With a three-finger robot having 7DoF, Okada et. al. [12] was able to detect the size of basic geometrical-shape objects with more than 60% accuracy and 3D shapes with more than 95% accuracy. By reconstructing the hand posture when about to grasp an object, using video frames of a pair of stereo cameras, Yizhou et al. [9] showed that object recognition can be improved by observing the shape of the hand from an egocentric point of view. Castellini et. al. [1] showed that the accuracy of the vision-based object recognition can be improved by combining hand posture when grasping the object. Our work extends these works by just focusing on the hand interactions when using an object, as we observe the motion and the posture to recognize the activity performed on an object.

3 *O*SENSE

Concept: Based on our observations, most of the daily used objects are handled in different ways, as those objects are used for specific tasks. At the same time, some objects associate similar activities when used. For example, the activity of drinking from a bottle vs. a mug has similar motions. At the same time, we grasp these objects in different ways as well: different hand postures in holding a bottle vs. a mug. However, similar postures occur when handling different objects, such as holding a knife vs. a hammer. To avoid confusion in independent use of activity or posture, with *O*Sense, we try to resolve confusions in action-based activity recognition by further using posture data in order to identify the object.

*Hardware: O*Sense was implemented using six MPU-9250 IMU sensors which contain a 3-axis accelerometer, 3-axis gyro and a 3-axis magnetometer, totalling 9 DoF. To gather data from IMUs, we used Teensy 3.6 development board with 32 bit 180MHz ARM Cortex-MX processor which has 3 easily accessible I2C ports. We made 5 rings with elastic bands, to be worn on the fingertips of the five fingers and on the back of the hand, each of which contained an IMU (Figure 1: RE). Each IMU was connected to the Teensy, worn on the wrist, and connected to the computer through a USB cable. In each sensing cycle, a total of 54 readings were obtained from the device: (3 accel + 3 gyro + 3 magneto) $\times$ 6 sensors.

Software: We created an Arduino program to read data from all 6 IMUs and compose it to a single data frame. A Java program was created to read the data frame from the device. To calculate posture of the hand, we implemented Madgwick quaternion update algorithm [10]. 35Hz was selected as the sample rate as to provide enough time to calculate quaternions without data loss.

4 EVALUATION

Study Design

Participants: We evaluated the system with 12 right-handed participants: 9 males and 3 females aged between 21 to 31 (mean = 26.3, SD = 3.4).

Activity Set: 9 different day-to-day common activities were chosen, associated with 3 different scenarios: workshop, kitchen and office (Figure 1). These activities associate with passive objects. *No-Action* was used as the baseline.

Procedure: The magnetometer of the device is calibrated first before each experiment. Then participants were asked to wear the device on the right hand and keep the hand still on a reference position for 20 seconds. This ensured the calibration is done and the device is stabilized. The experiment had three sessions, and in each session, nine activities had to be performed. Each activity took 55 seconds, during which 2000 data samples were collected. *No-Action* was recorded for 5 seconds between each activity. Activities were randomized in each session. Between sessions, participants had to take

Table 1: Accuracy comparison for different machine learning algorithms with different window sizes

Algorithm	NN				RF				SVM			
Window Size	50	100	150	200	50	100	150	200	50	100	150	200
Accuracy (%)	**71**	**72.75**	**74**	**74.75**	68	70.25	71.75	71.75	68	70	71.5	71

out the device from hand and wear it again. A session lasted around 13 mins and the entire experiment took about 40 mins per participant.

Data Analysis

Pre-Processing: Based on previous work [8], we only considered accelerometer data for the activity recognition. Performing the selected activities, there is little or no finger movement relative to the back of the hand. Hence we considered IMU on the back of the hand as the reference and only used its accelerometer data.

We calculated the total acceleration (L2 norm of the 3D acceleration vector) to make it a single scalar. DC component was removed when analyzing in the frequency domain. Using a moving window, eight statistical features were calculated: *mean*, *std*, *min*, *max*, *median*, *skewness*, *kurtosis*, *rms*. Then we applied Hamming window on the same data window and used Fast Fourier Transform (FFT) to calculate 17 more features in the frequency domain: first 5 top dominant frequencies, total power of the frequencies between 0.3Hz and 10Hz, power to total power ratios of 5 power values, ratio between the current dominant freq vs. previous.

To classify the posture, we followed the previous work that used joint angle measurement [14] and calculated the relative quaternion angles. Similar to acceleration measurement, we took back of the hand as the reference point and calculated angles between fingertips and the reference. Additionally, we calculated the angles between each finger as it gives more information about the posture. In total, there were 15 angles.

Classification Algorithms: For the activity classification, we compared 3 machine learning algorithms, mostly used in HAR tasks: Support Vector Machine(SVM), Random Forest(RF) and Neural Networks(NN). Four-fold cross validation was used to test the accuracy. We used the Python scikit-learn toolkit[1] for the algorithm implementation. While RF and SVM had default parameters, the settings for NN are: hidden_layer_size = 100×100, max_iter = $200,000$ and learning_rate = adaptive were used besides default parameters.

Results

Moving Window Size: We analyzed 4 sizes of windows with 90% overlapping using all 3 classification algorithms. As shown in Table 1, for every window size, NN provides the best accuracy and the higher the window size, the accuracy

DB	HA	CH	CK	DM	RE	WR	SA	SC	ST	
58.4	0.3	0.1	0.4	37.4	1.3	0.1	0.8	0.9	0.2	DB: Drinking (Bottle)
0.3	88.7	5.1	3.0	0.0	0.0	0.2	0.4	1.1	1.1	HA: Hammering
0.2	7.4	75.8	1.2	0.1	0.0	0.0	7.4	7.8	0.1	CH: Chopping
0.3	2.5	1.5	62.9	0.9	0.2	6.2	8.5	8.9	8.2	CK: Cutting (Knife)
41.6	0.0	0.1	0.6	54.8	1.2	0.1	0.7	0.6	0.3	DM: Drinking (Mug)
4.4	0.2	0.0	1.7	2.8	76.6	10.7	0.1	0.8	2.7	RE: Relaxed
0.2	0.3	0.0	5.3	0.2	3.2	88.0	0.0	0.9	1.8	WR: Writing
1.7	0.6	6.1	9.3	0.8	0.1	0.0	74.3	6.3	0.5	SA: Sawing
1.0	1.3	4.2	11.3	0.7	0.2	1.1	4.2	72.6	3.6	SC: Driving a Screw
0.1	0.9	0.0	10.0	0.3	1.0	2.8	2.1	3.4	79.4	ST: Stirring

Figure 2: Confusion matrix of activity classification for all participants (NN with window size of 100)

increases. The window size should be chosen based on the application, as large window size (i.e. more data) requires more time for classification.

Activity Classification: We ran NN over all users' data with a window size of 100. As shown in Figure 2, most activities (e.g. Hammering, Writing etc.) can be classified reasonably well, whereas two activities, drinking(mug) and drinking(bottle), gets confused.

Per-User Classification: We further analyzed the data to determine accuracy for personalized activity recognition. The results are shown in Figure 3-"before". As expected, personalized classification has higher accuracy compared to generic model. However, it has the same confusion between drink(bottle) and drink(mug).

Improving Object Recognition by Using Posture: As the posture provides hints about the object being grasped, we used it to resolve activity conflicts. During the posture classification model building, postures that show higher classification accuracy across users are selected and during the inference, if those postures have high accuracy (>95%), then the activity recognition results are filtered by them. If the activity recognition is low (<60%) again the posture prediction is considered. In all other scenarios, activity classification is used without changes, hence the ambiguities are reduced. Figure 3 shows the per-person accuracy of activity recognition before resolving the drink(bottle) and drink(mug) confusion and after resolving using posture data. Based on the t-test: $t(12) = -2.91$, $p \leq .05$, it is clear that the use of posture prediction resulted in significantly higher accuracy.

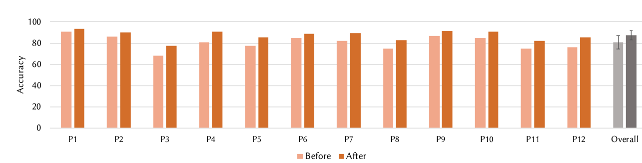

Figure 3: Per-person accuracy for activity classification before and after resolving conflicts with posture data (NN with window size of 100)

[1]https://scikit-learn.org/stable/

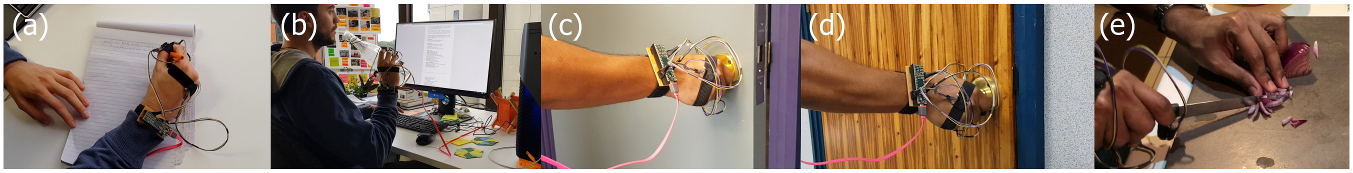

Figure 4: Example application scenarios (from left to right): a) track writing hours, b) track hydration level, c) Opening the home door in the morning, d) Opening the office door in the afternoon, and e) preparing a meal

5 APPLICATION SCENARIOS

Life-logging: While camera systems commonly used for life-logging applications, they have inherent privacy issues. *O*Sense alleviates this issue by avoiding using the cameras. *O*Sense can be used in scenarios such as the amount of time a person spends on writing each day to monitor how efficient the day was and how many times a person drank water from the bottle while at work to track the hydration level (Figure 4 a,b).

Just-in-time Information Applications: We can combine the output of *O*sense with time and location provide rich just-in-time information. For example, turning the door knob could be detected by the *O*Sense and if it is morning and at home, it could pop up the weather forecast, and schedule of the day. If it is the office door and it is evening, the system can pop up the grocery list to be bought on the way home (Figure 4 c,d).

Scaffolding Tasks: With the *O*Sense it is possible to count the number of repetitions and the activity duration. Hence it can be used to track the usage of a certain device and the duration of usage. For example, when preparing a meal, the user can be guided along the recipe: first cut the onion, then put into the mixer, and lastly stir it for 30 seconds (Figure 4e).

Hand Rehabilitation: Existing methods for hand rehabilitation include physical therapy(labour intensive) and glove based systems(lose the natural sensation of fingers). To improve hand functions, patients have to perform repetitive tasks to improve hand strength, accuracy, and range of motion [13], typically at a rehabilitation centre. *O*Sense can be used to track the rehabilitation progress at home as it can track the hand and finger movements.

6 LIMITATIONS AND FUTURE WORK

Object Detection: In this study we focused on classifying activities associated with particular objects. Hence our approach cannot be used to identify objects when held in static postures.

System Activation: There can be special cases where the users is not grasping any object, but the hand posture is similar to grasping some objects with a similar hand motion. To deal with such cases, a pressure sensor can be used so that the recognition algorithm will only run when there is an actual object in the hand.

IMU Calibration Issues and Number of IMUs: Before each session of the experiment, *O*Sense needed to be calibrated in order to account for drifts in sensor readings. Also since the magnetometer is highly susceptible to electromagnetic signals in the environment, posture data can be noisy. Hence future studies should focus on how to minimize these impacts or use different sensors for posture analysis which are invariant of the environment. Also to deploy *O*Sense in real world scenarios, the number of IMUs used should be reduced.

Expanding the number of Objects: When we selected objects for this study, we focused on a small set of passive objects in three scenarios: workshop, kitchen and office desk. To create a more robust system, we need to expand the number of objects in different scenarios.

REFERENCES

[1] Tommasi T. Noceti N. Odone F. Caputo B. Castellini, C. 2011. Using object affordances to improve object recognition. *IEEE Transactions on Autonomous Mental Development* 3, 3 (2011), 207–215.

[2] Umberto Castiello. 2005. The neuroscience of grasping. *Nature Reviews Neuroscience* 6, 9 (2005), 726.

[3] Chen S. Xie P. Chang C. Sun M. Chan, C. 2016. Recognition from hand cameras: A revisit with deep learning. In *European Conference on Computer Vision*. Springer, 505–521.

[4] Tapia E. M. Sadi S. Feldman, A. and Schmandt C. Maes, P. 2005. ReachMedia: On-the-move interaction with everyday objects. In *Ninth IEEE Int. Symposium on Wearable Computers (ISWC'05)*. IEEE, 52–59.

[5] Alexandra Grosskopf and Johann P Kuhtz-Buschbeck. 2006. Grasping with the left and right hand: a kinematic study. *Experimental Brain Research* 168, 1-2 (2006), 230–240.

[6] Xiao R. Harrison C. Laput, G. 2016. Viband: High-fidelity bio-acoustic sensing using commodity smartwatch accelerometers. In *Proc. UIST*. ACM, 321–333.

[7] Yang C. Xiao R. Sample A. Harrison C. Laput, G. 2015. Em-sense: Touch recognition of uninstrumented, electrical and electromechanical objects. In *Proc. UIST*. ACM, 157–166.

[8] Labrador M. A. Lara, O. et al. 2013. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials* 15, 3 (2013), 1192–1209.

[9] Hua G. Mordohai P. Lin, Y. 2014. Egocentric object recognition leveraging the 3d shape of the grasping hand. In *European Conference on Computer Vision*. Springer, 746–762.

[10] Harrison A. Vaidyanathan R. Madgwick, S. 2011. Estimation of IMU and MARG orientation using a gradient descent algorithm. In *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*. IEEE, 1–7.

[11] Kishino Y. Sakurai Y. Suyama T. Maekawa, T. 2011. Recognizing the use of portable electrical devices with hand-worn magnetic sensors. In *International Conference on Pervasive Computing*. Springer, 276–293.

[12] T Okada and S Tsuchiya. 1977. Object recognition by grasping. *Pattern Recognition* 9, 3 (1977), 111–119.

[13] Lyne S. Wang Z. Nicolini L. Mosadegh B. Whitesides G. Walsh C. Polygerinos, P. 2013. Towards a soft pneumatic glove for hand rehabilitation. In *IROS, 2013 IEEE/RSJ Int. Con. on.* IEEE, 1512–1517.

[14] Luinge H. Slycke P. Roetenberg, D. 2009. Xsens MVN: full 6DOF human motion tracking using miniature inertial sensors. *Xsens Motion Technologies BV, Tech. Rep* 1 (2009).

[15] Adamovich S. Poizner H. Schettino, L. F. 2003. Effects of object shape and visual feedback on hand configuration during grasping. *Experimental Brain Research* 151, 2 (2003), 158–166.

[16] Grossman T. Wigdor D. Fitzmaurice G. Yang, X. 2012. Magic finger: always-available input through finger instrumentation. In *Proc. UIST*. ACM, 147–156.