Towards Digital Apprenticeship: Wearable Activity Recognition in the Workshop Setting

Timothy Campbell
Jonathan Harper
Björn Hartmann
Eric Paulos

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2015-172
http://www.eecs.berkeley.edu/Pubs/TechRpts/2015/EECS-2015-172.html

July 15, 2015

Towards Digital Apprenticeship: Wearable Activity Recognition in the Workshop Setting

Tim Campbell, Jonathan Harper, Björn Hartmann, Eric Paulos
University of California, Berkeley
{timcam, jharper, bjoern, paulos}@berkeley.edu

ABSTRACT
Online communities for sharing instructional content have grown from a renewed interest in DIY culture. However, it is difficult to convey the tacit knowledge implicit in certain skills. We identify the need for *Digital Apprenticeship*, where workshop activities are sensed and analyzed for both quantitative and qualitative measures. We evaluated this concept with an activity recognition system for carpentry tools. Using a single ring-worn inertial measurement unit (IMU), we collected data from 15 participants using 5 hand and power tools. Our window-based multi-class SVM achieves 82% accuracy with realistic training scenario and outputs user-friendly event activity. We investigate how these results contextualize to applications in digital apprenticeship, namely tutorial authoring, content following and technique feedback.

Author Keywords
digital fabrication; DIY; activity recognition; tutorials; wearable computing; inertial measurement unit (IMU)

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
The Maker Movement has revived interest in manual fabrication domains like carpentry and welding. In lieu of formal training, amateur craftsmen often turn to online tutorial sites like Instructables[1] and Make[2] for instructional content and craft knowledge [15]. Though well-written tutorials may be an effective way for online communities to share some skills, static content is not well suited for teaching physical skills that involve tacit knowledge.

Often taught under supervision of an expert, tacit knowledge is the subtle, nuanced skill that appears in almost every physical craft yet goes unrecorded in documentation. We aim to reproduce a master-apprentice relationship through dynamic tutorials that provide real-time feedback to the user. A wearable sensor could be used to record workshop activity and generate step-by-step instructions; similarly, users could follow and gain feedback from an existing tutorial. We term this vision *Digital Apprenticeship* - where amateur users are instructed with specific corrections on not only *what* tool they use, but also on *how* they use that tool.

This note explores how dynamic tutorials can support distributed learning. We show that an activity recognition system can supply tutorial applications with usable, human-readable events. Training data were collected in a controlled workshop experiment using a only a ring-worn wireless sensor. Fifteen participants performed carpentry tasks with 5 common hand and power tools. Our inertial sensor streams 9 degree-of-freedom (DOF) data: accelerometer, gyroscope, and magnetometer. We built a window-based classifier using a multi-class SVM to identify discrete time periods of tool use (e.g. "hammering for 10 seconds"). The system was evaluated using a leave-one-user-out validation scheme, meaning new users can achieve the reported accuracy without providing any training data. We report 82% accuracy.

This work builds on previous studies by Ward [16, 17] and Lukowicz [11] by developing a separate approach to classifying workshop activity. Our system generates human-readable output events, which we apply to three target applications: tutorial authoring, tutorial following, and technique feedback. Using characteristic results, we show how two common errors affect the tutorial authoring and tutorial following applications. Finally, we demonstrate qualitative feedback by classifying three distinct types of sanding where an interface could guide the user towards a specific technique.

RELATED WORK
Craft communities have turned to online sharing platforms to share projects and skills. In response, HCI research has begun to consider the effectiveness of sharing physical craft making online [2]. Lindtner et. al. further argues that makerspaces are blending both amateur and professional development, especially in areas of HCI innovation [10]. Using sensing and activity recognition, we see an opportunity to build sharing tools for physical skills.

Human activity can often be reconstructed from a few inertial sensors [8]. Placing IMUs near the activity source enables more accurate activity recognition. Foxlin et. al. showed pedestrian tracking on the shoe [5] and the RecoFit system

[1] `http://www.instructables.com`
[2] `http://www.makezine.com/projects/`

Submitted as a technical report for UC Berkeley EECS, July 2015.

identifies exercises from an arm-worn IMU [12]. Sensing physical fabrication skills requires that our sensing device be unobtrusive; our device enables 9 degree-of-freedom (DOF) inertial sensing on the finger [6].

Activity Recognition for Assembly Tasks

Researchers have developed techniques for tracking assembly tasks for toys, furniture, and manufacturing. DuploTrack [7] uses the Microsoft Kinect to automatically generate assembly instructions for Duplo blocks, yet it is unknown whether video scales to complex assembly tasks. Other work has applied activity recognition to furniture assembly [1] or automotive manufacturing tasks [14]. However, these systems are designed to track construction against a predefined set of instructions.

Activity Recognition in Manual Fabrication

Previous work used a combination of microphones and accelerometers in both wearable and distributed formats with experiments performed in a wood workshop [11, 17]. Ward et. al. used a subset of collected data to classify similar workshop actions with a wrist-worn accelerometer and microphone [16]. The authors achieved 70.1% classification on 9 tools using a within-subjects training scheme.[3] In group settings such as hackerspaces with high ambient noise, microphones may be unsuitable. Our approach focused on a single IMU that leverages the unique physical and magnetic properties of tools in these settings.

SYSTEM DESIGN & IMPLEMENTATION

Our system classifies workshop activity from time-series data and then outputs whole-event activities. The data stream is buffered into windows and classified as independent events. Final events are gathered from the single window classifications using a smoothing convolution, which produces a set of labeled time windows. To train the classifier we conducted a data collection experiment with 5 workshop tools.

Hardware

We employ a finger-worn sensing platform built atop the GINA Mote [6]. The device is configured as a wireless inertial measurement unit (IMU) with a 3-axis accelerometer, gyroscope, and magnetometer for a total of 9 degrees-of-freedom. The accelerometer sensitivity is set to +/- 6 G. The gyroscope senses rotational velocity at a maximum rate of 2000 Degrees per second. Finally, the magnetometer measures orientation and also detects fluctuating magnetic fields with a +/- 4 Gauss range. Processing on the ring is based on the Texas Instruments MSP430 chipset, which is a 16-bit, 16 MHz microprocessor.

Data are streamed directly from the ring to the computer via an Atmel 2.4 GHz transceiver at 150 packets per second, though this limits battery life to approximately 20 minutes. The circuit board has a footprint of 12 x 15 mm and the entire device fits within a 3D printed ring.

[3]For comparison, we achieved 86% accuracy across 5 tools when our classifier trained within-subjects. Throughout this paper we train on leave-one-user-out scenario and report 82% accuracy.

Methods

We collected data through scripted use of 5 tools (*hammer, cordless drill, hand driver, saw, and power saw*) with multiple repetitions. Each participant was asked to complete the a series of unrelated tasks (e.g. drill 10 holes) with their most comfortable technique. Participants wore the ring on their dominant hand as shown in Figure 1, and each participant confirmed that the ring did not impede their activity. Data collection for each participant took 30 minutes, which we recorded in a series of trials. Also, we recorded video and audio of each trial for annotation purposes. A total of 15 participants (10 males, 5 females) completed the data collection.

Annotation

Since our classifer requires labeled training data, we built an annotation interface to manually translate the video with respect to sensor data. We annotated whole activities (e.g. *"hammering"*) to establish a ground truth comparable to the desired output and the interface outputs a set of labeled time windows. Table 1 shows the annotated events for each tool. We balanced the protocol to contain approximately equal repetitions of each tool, shown in the Window Count row.

Data Analysis

We base our method for classifying workshop activity on electromyography research by Saponas et. al. [13]. The researchers divided a time-series signal into windows, classified the windows independently and then took a majority vote to determine whole event classification. Windowing the data gives the classifier a time-independent signal.

Signal Processing

We first divide the data into windows by discrete time intervals of 400 *ms*. We set the window size empirically based on the results of cross-validation. To remove variation in the received packet rate, the data are resampled to 150 Hz. Each window contains 9 vectors of IMU data from the 3 sensors (accelerometer, gyroscope, and magnetometer). Once the data has been windowed, labels are attached to each window only if the window falls entirely within the time range of the annotation. Otherwise, the window is labeled as noise. Table 1 shows the computed number of windows for each tool. Notice that short actions like drill have proportionally fewer windows than other tools.

Machine Learning Classifier

We calculate a total of 144 features on the samples in each window of data. *Root Mean Square* (RMS) gives a measure

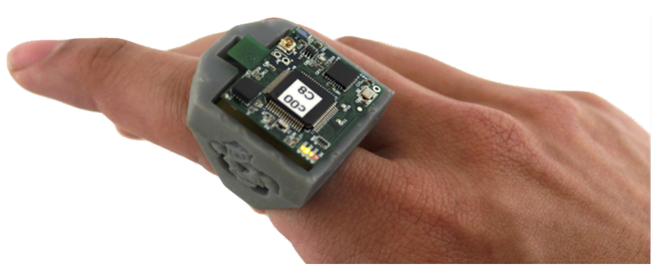

Figure 1. The wireless GINA Mote based wireless 9 DOF inertial sensing hardware as worn by a user. Experiment participants placed the ring on the index or middle finger of their dominant hand.

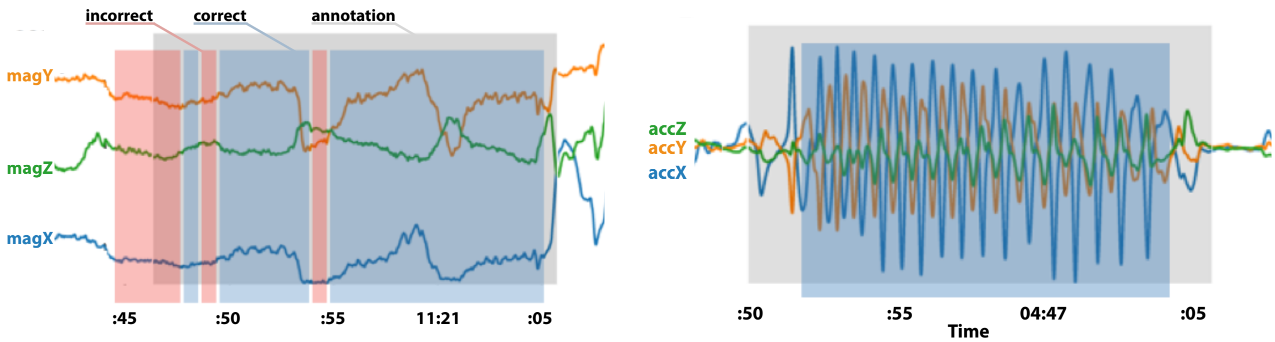

Figure 2. The results of two trials are shown with annotated events in gray, correct classifications are in blue and incorrect in red. (a) A screwdriver trial with magnetometer signal achieved 81% classification yes shows how the output can "bounce" between two classes. (b) A saw trial with accelerometer signal shown gave a similar accuracy of 78% and only missed the start and stop sections of the event.

of the energy for each of the 9 signals. We then take ratios of RMS features with respect to the other 8 features, exclusive of repeated ratios, for another 36 features. Finally, the *Variance* of each signal adds 9 features.

Frequency Energy is computed for each sensor. Using cross-validation we found 30 frequency bins as optimal for our domain. With 120 to 150 Hz sampling frequency, this limits the Nyquist frequency to approximately 60 Hz, or 2 Hz per bin. We compute the fast-fourier transform (FFT) of all 9 data vectors independently then sum the three axes for each sensor into a vector of length 30. This removes orientation effects. The final result is 90 FFT features per window.

In our initial experiments, we tried several methods for classification and found that Support Vector Machines (SVMs) performed well for classifying independent windows. We implemented our classifier in MATLAB using the **liblinear** [4] package for its multi-class SVM functionality. Multi-class SVMs operate similar to single-class SVMs except a hyperplane is constructed for each class against all other classes. After classifying single windows, the system passes a uniform smoothing kernel over the predictions to remove erroneous windows from an otherwise uniform prediction. Finally the system groups the smoothed, consecutive windows into time windows with activity labels, shown in Figure 2b.

Classification Results

Accuracy on single-window classification is measured as the percent of windows classified correctly against our ground truth labels. We tested the classifier using leave-one-user-out cross-validation and achieved an average accuracy of 82.44% across the 6 classes and 15 users. Figure 3 shows the confusion matrix for our classifier. Unsurprisingly, noise is frequently misclassified for other tools because it often contains features that match other tools. The drill had the lowest true positive rate; we believe this is due to relatively short activities. Removing the drill class increases the accuracy to 89.07%. Figure 2 highlights two common results of our activity recognition system: "bouncing" between states and start/end differences. Both trials have single-window accuracy around our reported average, but vary in their usability.

Tool	Drill	Driver	Hammer	Saw	Skilsaw	Noise	Total
Annotated Events (#)	179	73	106	70	74	0	502
Window Count (#)	1548	1794	1870	2320	1325	6542	15399

Table 1. Tabulations for events and windows for each tool. Participants completed over 500 workshop activities (i.e. hammering a nail).

APPLICATIONS

Our activity recognition system successfully takes continuous time-series data and extracts meaningful events. To explore digital apprenticeship, we apply our output to three key areas: *automatic content publishing, quantitative step-by-step feedback, and qualitative skill-based feedback*. These applications provide insight into designing mixed-initiative insterfaces, especially when erroneous classifications cause the user extra effort [9].

Tutorial Authoring

We used existing tutorial editing software, DemoCut [3], that takes user video, asks for manual tags and produces a step-by-step tutorial. Combining our system with DemoCut, we uploaded trial videos and our system's output. DemoCut successfully produced tutorial videos optimized for online sharing as shown in Figure 4a. With correctly recognized activities, user intervention is only necessary to adjust the start and end times, a common error shown in Figure 2a. However for a trial that bounces between states, as in Figure 2b, the added work to remove erroneous classifications may outweigh the benefit of automatic tags in the first place.

Figure 3. Confusion matrix for the single window classifications. Each row represents ground truth labels. Overall accuracy was 82.44%

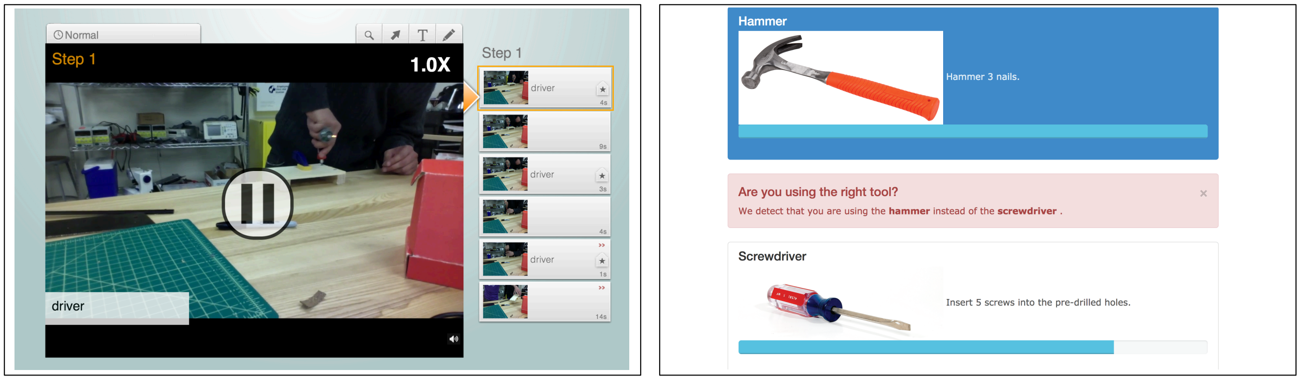

Figure 4. Two target application interfaces. (a) A video annotation interface with plays short clips of each automatically generated tag. (b) A tutorial following interface showing completed steps and correction on tool usage.

Tutorial Following

To demonstrate tutorial following, we prototyped a web interface that tracks the completion of each step in a tutorial. When the interface detects a tool is used out of order, it corrects the user towards the current task. We built and evaluated our system offline, though window-based classifiers also work in real-time recognition. Thus, we streamed the results of our activity recognition into the interface prototype in real-time. In the best case, tutorial following allows craftsmen to follow step-by-step instructions without interruption. However, when we feed in misclassified data, the interface cannot track progress and prompts excessive corrections, even though the user knows she has the correct tool.

Qualitative Feedback for Physical Skills

We demonstrate a form of qualitative feedback with sanding, an essential skill in carpentry. We recorded 4 trials of sanding, each containing 3 classes: *straight*, *small circles*, and *large circles*. Without modifying our system we ran leave-one-out training and achieved 85% accuracy on detecting sanding skill. Thus, our system could correct suggest an improved technique for the given task (e.g. "try sanding in a circular pattern"). Expert craftsmen could record themselves performing skills along a scale of good and bad technique to convey deeper tacit knowledge.

DISCUSSION

In order to support the vision of Digital Apprenticeship, we focused on a system that required minimal instrumentation and tailored the results to tutorial applications. We also prototyped a system to capture tacit knowledge and showed initial results by partitioning the sanding technique into three distinct classes. Here we discuss how Digital Apprenticeship supports a wider range of interactions that focus on providing feedback to users in the workshop.

Broader skilled labor

Activity tracking in the workshop setting also extends to the skilled labor force [10]. The Proglove[4] provides feedback to skilled workers on the assembly line, but relies on instrumented tools outfitted with RFID for sensing and adherence. An IMU-centric approach leverages the unique inertial characteristics of tools without the need to install infrastructure in the environment, for more impromptu interactions. Skilled laborers often perform tasks that automation cannot, like completing a surface finish. These qualitative actions are essential for future activity tracking especially as smart tools enter the design and fabrication environment [18].

Mixed-initiative interfaces

The applications we showed can passively offer suggestions but raise issues of mixed-initiative interfaces [9]. These intelligent agents provide some form of automation and could offer feedback based on the confidence of the classification. For instance, a feedback system could detect unsafe activity and prevent injury, triggering the a smart tool to power down. However, we showed how the "bouncing" output events can also make for improper feedback. In full-featured systems, these small misclassifications should be accounted for.

Limitations and Future Work

With a constrained dataset, this work does not explore a full project-based workshop activity, especially using tools that may not be suited for inertial sensing (e.g. soldering, glueing). Future studies will expand data collection to a larger suite of workshop tools. Some features are not well recognized by our system; for example, our naive windowing approach misses the magnetic features found at the start of drill usage. Future work will more closely consider unique physical properties of each tool, and develop signal processing routines which could attempt to first recognize motors or repetitive motions before feeding into a classifier.

Conclusion

In this paper we detailed activity recognition for workshop activities using a single, ring-worn IMU. We built a window-based classifier for continuous time-series data and smoothed the output into user-friendly events. Our leave-one-user-out training represents a realistic user scenario and we make accuracy improvements over previous work. Towards *Digital Apprenticeship*, we applied the classified events to tutorial authoring and following applications that captured both quantitative user activity and also qualitative user technique.

[4] `http://www.proglove.de/`

REFERENCES

1. Antifakos, S., Michahelles, F., and Schiele, B. Proactive instructions for furniture assembly. In *Ubiquitous Computing*. Springer, 2002, 351–360.
2. Buechley, L., Rosner, D. K., Paulos, E., and Williams, A. DIY for CHI: Methods, Communities, and Values of Reuse and Customization. In *Proc. of CHI Extended Abstracts*, ACM (2009), 4823–4826.
3. Chi, P.-Y., Liu, J., Linder, J., Dontcheva, M., Li, W., and Hartmann, B. DemoCut: Generating Concise Instructional Videos for Physical Demonstrations. In *Proc. of UIST* (2013), 141–150.
4. Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J. LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research 9* (2008), 1871–1874.
5. Foxlin, E. Pedestrian tracking with shoe-mounted inertial sensors. *IEEE Computer Graphics and Applications 25*, 6 (2005), 38–46.
6. Greenspun, J., and Pister, K. S. Ring GINA: a wearable computer interaction device. In *Mobile Computing, Applications, and Services*. Springer, 2014, 98–103.
7. Gupta, A., Fox, D., Curless, B., and Cohen, M. DuploTrack: a real-time system for authoring and guiding duplo block assembly. In *Proc. of UIST* (2012), 389–402.
8. Harrison, B. L., Consolvo, S., and Choudhury, T. Using multi-modal sensing for human activity modeling in the real world. In *Handbook of Ambient Intelligence and Smart Environments*. Springer, 2010, 463–478.
9. Horvitz, E. Principles of mixed-initiative user interfaces. In *Proc. of CHI* (1999), 159–166.
10. Lindtner, S., Hertz, G. D., and Dourish, P. Emerging Sites of HCI Innovation: Hackerspaces, Hardware Startups & Incubators. In *Proc. of CHI*, ACM (2014), 439–448.
11. Lukowicz, P., Ward, J. A., Junker, H., Stger, M., Trster, G., Atrash, A., and Starner, T. Recognizing workshop activity using body worn microphones and accelerometers. In *Pervasive Computing*. Springer, 2004, 18–32.
12. Morris, D., Saponas, T. S., Guillory, A., and Kelner, I. RecoFit: using a wearable sensor to find, recognize, and count repetitive exercises. In *Proc. of CHI* (2014), 3225–3234.
13. Saponas, T. S., Tan, D. S., Morris, D., and Balakrishnan, R. Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces. In *Proc. of CHI* (2008), 515–524.
14. Stiefmeier, T., Roggen, D., Ogris, G., Lukowicz, P., and Tröster, G. Wearable activity tracking in car manufacturing. *IEEE Pervasive Computing 7*, 2 (2008).
15. Torrey, C., McDonald, D. W., Schilit, B. N., and Bly, S. How-to pages: Informal systems of expertise sharing. In *ECSCW*. Springer, 2007, 391–410.
16. Ward, J., Lukowicz, P., Troster, G., and Starner, T. Activity Recognition of Assembly Tasks Using Body-Worn Microphones and Accelerometers. *IEEE Transactions on Pattern Analysis and Machine Intelligence 28*, 10 (Oct. 2006), 1553–1567.
17. Ward, J. A., Lukowicz, P., and Trster, G. Gesture spotting using wrist worn microphone and 3-axis accelerometer. In *Proc. of Smart Objects and Ambient Intelligence*, ACM (2005), 99–104.
18. Weichel, C., Alexander, J., Karnik, A., and Gellersen, H. SPATA: Spatio-Tangible Tools for Fabrication-Aware Design. In *Proc. of TEI*, ACM (2015), 189–196.