A Demonstration of Inferring Identity using Accelerometers in Television Remote Controls

Jeffrey Hightower∗, Branislav Kveton†, and Keng-hao Chang‡

∗Intel Research, Seattle, Washington, USA
†Intel Research, Santa Clara, California, USA
‡University of California, Berkeley, California, USA

Abstract. This demonstration shows how accelerometers and button-press loggers embedded in a television remote control can be used to distinguish different participants based on the unique way each person wields the remote. This is a live demonstration of the inference technology presented in the paper *Inferring Identity using Accelerometers in Television Remote Controls*, which appears as a full paper in Pervasive 2009. Unlike the full paper, which presents results from five 3-week data sets collected from real homes, this demonstration simplifies the machine learning and instead does user-classification with real-time training data collected from visitors to the demo kiosk.

1 Overview

We demonstrate a new lightweight biometric that analyzes participants' hand motions and button press sequences. Specifically, we show that accelerometers and and button-press logger embedded in a television remote control can be used to distinguish different participants based on the unique ways each person wields the remote. This demonstration is a simplified version of the technique presented and evaluated in *Inferring Identity using Accelerometers in Television Remote Controls*, which appears in the Pervasive 2009 full paper track [1].

The goal of identity inference in this application is to personalize the television watching experience so that service providers, content creators, and consumer electronics manufacturers can expand their user-base, provide exciting and relevant programming, increase the effectiveness of advertising [2], incorporate digital home technologies like interactive TV [3], and distinguish their devices' features and usability. For example, a digital video recorder manufacturer could incorporate this capability to allow its devices to provide program recommendations per user instead of per device. The demonstration also shows how machine learning can be made simple enough to be invisible and embedded a in pervasive computing application. Users can simply grasp the remote control and watch TV without any effort to explicitly login or identify themselves. Our system observes a participant's hand motion in the background and analyzes whether it matches existing identity signatures.

2 Demonstration Kiosk

Figure 1 shows the hardware and software components of our demonstration kiosk. The demonstration itself requires only a table and the ability to plug the laptop into the power mains.

Fig. 1: Data Collection Components: top, laptop receives sensor streams, trains and runs the classifier, and graphically displays the results; bottom left to right, video camera captures a live view of the room from the television's perspective, remote control with attached accelerometer module attached, universal infrared code receiver.

Accelerometer Module Our 3-axis accelerometer module is attached and wired into the power source of a TiVo$^{\mathrm{TM}}$ remote control. The accelerometer module continuously measures and transmits all the acceleration forces imposed on the remote control. The module hardware is a custom 3-axis accelerometer board connected to a Telos sensor mote [4], which acts as a relay to transmit the data to the laptop. The module is enclosed in a custom plastic case.

Infrared Receiver We use the Tira-2.1 multi-protocol infrared receiver made by HomeElectronics. This receiver sits by the TV to capture the infrared signals sent by button presses on the remote control. Each button press is timestamped and logged by its unique ASCII code string.

Laptop Computer The laptop receives and processes the acceleration and button press data streams. Acceleration data is wirelessly transmitted to the laptop using another paired Telos mote plugged into the laptop's USB port. The infrared code stream is received through a direct USB connection to the Tira infrared receiver. All data is timestamped with 100ns precision. The laptop runs the classification algorithms and presents a graphical display to guide demo participants through the training and recognition steps.

Video Camera The last component is a video camera pointed at the room where the TV is located. In collecting data for the full paper the camera was used to capture hand-coded ground-truth information about who was watching TV, however the demonstration simply uses the camera to present participants with a TV-centric viewpoint of themselves and the environment.

2.1 Demo Participants' Experience

The demo participants' experience can be broken into three stages, which correspond to a typical supervised machine learning pipeline.

Training The participant is prompted to do several actions characteristic of watching television, specifically turn on the television, surf between several channels, enter 4 channels directly, manipulate the volume controls, and swap between prior channels. Acceleration and button-press sensor data is logged and labeled during this process and training can be repeated for up to 5 people. Five participants is not a functional limit—the inference algorithm scales arbitrarily—but demo pilot tests have suggested that exceeding five people makes it difficult for everyone to see the displays and participate fully.

Model Creation The classification algorithms are trained on the exemplars provided by each participant in the first training step. Computing the classifier model takes less than 1 second.

Identity Inference The system begins online recognition. Any participant may pick up the remote and press its buttons to simulate watching television. A classification decision is triggered by each button press on the remote. Based on the button-press and acceleration sensor data fed into the classifier, the laptop display shows who it believes to be the most likely participant using the remote along with the probability distribution of that estimate.

3 Classification Algorithm

The full paper on this work uses a graphical model (specifically, a Max-Margin Markov Network [5]) to build a long-term model that can look across several button presses and TV viewing sessions when estimating the participant's identity. In contrast, the classification algorithm for this demonstration is a straightforward Support Vector Machine (SVM) using a radial kernel with a gamma of $\frac{1}{\#ofparticipants}$ and degree 3. The SVM infers the identity of the participant triggered by each button press. Simplifying the machine learning in this way was necessary so we can train the demo system quickly to whoever walks up to the kiosk. The trade-off is obviously that the demo does not have the same generality and accuracy as the M^{3}N version. Our testing, however, has shown that the accuracy of the demo is in fact similar to the full system at around 80% because the demo does not need to handle as many behavior variations and corner cases as the full system deployed in someone's home.

The features for SVM training and classification are derived from the accelerometer and button press data stream. The full paper uses two-level extraction of 372 features including **button-press-level features** before, during, and after each button press and **session-level features** describing the relationships between several button presses in a sequence. The live demonstration uses only a 43 element subset of the button-press-level features. Given a button press at time t, we look at hand motion acceleration features during the preceding 0.5, 1, 2, and 4 seconds. For each window we compute the (1) energy, (2) dominant frequency, (3) magnitude of the fundamental frequency, (4) mean, (5) variance, (6) maximum, (7) minimum, (8) median (9) range, and (10) correlation coefficient. The first nine features are extracted for each of the x-, y-, and z-axis of the accelerometer. Energy, capturing the total amount of hand motion, is calculated by the squared sum of the results of a Fast Fourier Transform (FFT) with the DC component excluded. Fundamental frequency is the frequency with the highest magnitude from the result of FFT (again, with DC removed), which provides information about shaking. The three-axis mean serves as an indicator of the remote control's orientation. The correlation coefficient is extracted from each of the x-y, y-z, x-z axis pairs, calculated as $(\Sigma a_i b_i - \overline{a}\overline{b})/((n-1)\sigma_a \sigma_b)$ where a and b are sequences of n measurements with mean $\overline{a}$ and $\overline{b}$ and standard deviation σ_a and σ_b. Finally, we add features about the button press itself including (1) infrared code, (2) number of times the same code was sequentially transmitted, and (3) approximate duration of the button press.

4 Summary

This demonstration offers participants a "hands-on" experience with machine learning applied to a pervasive computing problem. It shows the technology we built to validate the hypothesis that accelerometers and key-press loggers embedded in a television remote control can be used to distinguish people based on the unique way each person wields the remote.

References

1. Chang, K., Hightower, J., Kveton, B.: Inferring identity using accelerometers in television remote controls. In: Proceedings of the Seventh International Conference on Pervasive Computing, Nara, Japan (May 2009)
2. von Rimscha, M.B., Rademacher, P., Thomas, N., Siegert, G.: The future of TV commercials. In: Annual Meeting of the International Communication Association, San Francisco, CA (May 23 2007)
3. Lu, K.Y.: Interaction design principles for interactive television. Master's thesis, Georgia Institute of Technology (May 2005)
4. Polastre, J., Szewczyk, R., , Culler, D.: Telos: Enabling ultra-low power wireless research. In: Information Processing in Sensor Networks, 2005. IPSN 2005. (April 2005) 364– 369
5. Taskar, B., Guestrin, C., Koller, D.: Max-margin Markov networks. In: Advances in Neural Information Processing Systems 16. (2004)