uWave: Accelerometer-based Personalized Gesture Recognition and Its Applications

Jiayang Liu, Zhen Wang, and Lin Zhong
Department. Of Electrical Computer Engineering
Rice University, Houston TX 77005
jiayang@rice.edu, wangzhen127@gmail.com, lzhong@rice.edu

Jehan Wickramasuriya and Venu Vasudevan
Pervasive Platforms & Architectures Lab
Applications & Software Research Center, Motorola Labs
{jehan,venu}@motorola.com

***Abstract*—The proliferation of accelerometers on consumer electronics has brought an opportunity for interaction based on gestures or physical manipulation of the devices. We present uWave, an efficient recognition algorithm for such interaction using a single three-axis accelerometer. Unlike statistical methods, uWave requires a single training sample for each gesture pattern and allows users to employ personalized gestures and physical manipulations. We evaluate uWave using a large gesture library with over 4000 samples collected from eight users over an elongated period of time for a gesture vocabulary with eight gesture patterns identified by a Nokia research. It shows that uWave achieves 98.6% accuracy, competitive with statistical methods that require significantly more training samples. Our evaluation data set is the largest and most extensive in published studies, to the best of our knowledge. We also present applications of uWave in gesture-based user authentication and interaction with three-dimensional mobile user interfaces using user created gestures.**

Keywords-gesture recognition, acceleration, dynamic time warping, personalized gesture

I. INTRODUCTION

Gestures[1] have recently become attractive for spontaneous interaction with consumer electronics and mobile devices in the context of pervasive computing [1-3]. However, there are multiple technical challenges to gesture-based interaction. Firstly, unlike many pattern recognition problems, e.g. speech recognition, gesture recognition lacks a standardized or widely accepted "vocabulary". It is often desirable and necessary for users to create their own gestures, or personalized gestures. With personalized gestures, it is difficult to collect a large set of training samples necessary for established statistical methods, e.g., Hidden Markov Model (HMM) [4-6]. Secondly, spontaneous interaction requires immediate engagement, i.e., the overhead of setting up the recognition instrumentation should be minimal. More importantly, the targeted platforms for personalized gesture recognition are usually highly constrained in cost and system resources, including battery, computing power, and interface hardware, e.g. buttons. As a result, computer vision [1, 2] or "glove" [3] based solutions are unsuitable.

In this work, we present uWave to address these challenges and focus on gestures without regard to finger movement, such as sign languages. Our goal is to support efficient personalized gesture recognition on a wide range of devices, in particular, on resource-constrained systems. Unlike statistical methods [4], uWave only requires a single training sample to start; unlike computer vision-based methods [7], uWave only employs a three-axis accelerometer that has already appeared in numerous consumer electronics, e.g. Nintendo Wii remote, and mobile device, e.g. Apple iPhone. uWave matches the accelerometer readings for an unknown gesture with those for a vocabulary of known gestures, or *templates*, based on dynamic time warping (DTW) [8]. uWave is efficient and thus amenable to implementation on resource-constrained platforms. We have implemented multiple prototypes of uWave on various platforms, including Smartphones, microcontroller, and the Nintendo Wii remote hardware [9]. Our measurement shows that uWave recognizes a gesture from an eight-gesture vocabulary in 2ms on a modern laptop, 4ms on a Pocket PC, and 300ms on a 16-bit microcontroller, without any complicated optimization.

We evaluate uWave with a gesture vocabulary identified by a Nokia research [4] for which we have collected a library of 4480 gestures from eight participants over multiple weeks. The evaluation shows that uWave achieves accuracy of 98.6% and 93.5% with and without template adaptation, respectively, for user-dependent gesture recognition. The accuracy is the best for accelerometer-based user-dependent gesture recognition. Moreover, our evaluation data set is also the largest and most extensive in published studies, to the best of our knowledge.

In summary, we make the following contributions.

- We present uWave, an efficient gesture recognition method based on a single accelerometer using dynamic time warping (DTW). uWave requires a single training sample per vocabulary gesture.
- We show that there are considerable variations in gestures collected over a long time and in gestures collected from multiple users; we highlight the importance of adaptive and user-dependent recognition.
- We report an extensive evaluation of uWave with over 4000 gesture samples collected from eight users over

[1] We use "gestures" to refer to free-space hand movements that physically move or disturb the interaction device. Such movements include not only gestures as we commonly know; but also physical manipulations like shaking and tapping of the device

multiple weeks for a predefined vocabulary of eight gesture patterns.

- We present two applications of uWave: gesture-based user authentication and gesture-based manipulation of three-dimensional user interfaces on mobile phones.

The strength of uWave in user-dependent gesture recognition makes it ideal for personalized gesture-based interaction. With uWave, users can create simple personal gestures for frequent interaction. Its simplicity, efficiency, and minimal hardware requirement (a single accelerometer) make uWave have the potential to enable personalized gesture-based interaction with a broad range of devices.

The rest of the paper is organized as follows. We discuss related work in Section II and then present the technical details of uWave in Section III. We next describe a prototype implementation of uWave using the Wii remote in Section IV. We report an evaluation of uWave through a large database for a predefined gesture vocabulary of eight simple gestures in Section V. We present the application of uWave to gesture-based user authentication and interaction with mobile phones in Section VI. We discuss the limitations of uWave and acceleration-based gesture recognition in general in Section VII and conclude in Section VIII. Two prototypes of uWave based on a Wii remote and a mobile phone, respectively, has been demonstrated at ACM UIST in October 2008 [10]. In this work, we present the technical details, system implementation, and applications of uWave.

II. Related Work

Gesture recognition has been extensively investigated [1, 2]. The majority of the past work has focused on detecting the contour of hand movement. Computer vision techniques in different forms have been extensively explored in this direction [7]. As a recent example, the Wii remote has a "camera" (IR sensor) inside the remote and detects motion by tracking the relative movement of IR transmitters mounted on the display. It basically translates a "gesture" into "handwriting", lending itself to a rich set of handwriting recognition techniques. Vision-based methods, however, are fundamentally limited by their hardware requirements (i.e. cameras or transmitters) and high computation load. Similarly, "smart glove" based solutions [3, 11, 12] can recognize very fine gestures, e.g., the finger movement and conformation but require the user to wear a glove tagged with multiple sensors to capture finger and hand motions in fine granularity. As a result, they are unfit for spontaneous interaction due to the high overhead of engagement.

As ultra low-power low-cost accelerometers appear on consumer electronics and mobile devices, many have recently investigated gesture recognition based on the time series of acceleration, often with additional information from a gyroscope or compass. Signal processing and ad hoc recognition methods were explored in [13, 14]. LiveMove Pro [15] from Ailive provides a gesture recognition library based on the accelerometer in the Wii remote. Unlike uWave, LiveMove Pro targets user-independent gesture recognition with a predefined gesture vocabulary and requires 5 to 10 training samples for each gesture. No systematic evaluation of the accuracy of LiveMove Pro is publicly available. HMM, investigated in [4, 5, 16], is the mainstream method for speech recognition. However, HMM-based methods require extensive training data to be effective. The authors of [5] realized this and attempted to address it by converting two samples into a large set of training data by adding Gaussian noise. While the authors showed improved accuracy, the effectiveness of this method is likely to be highly limited because it essentially assumes that variations in human gestures are Gaussian. In contrast, uWave requires as few as a single training sample for each gesture and delivers competitive accuracy. Another limitation of HMM-based methods is that they often require knowledge of the vocabulary in order to configure the models properly, e.g. the number of states in the model. Therefore, HMM-based methods may suffer when users are allowed to choose gestures freely, or for personalized gesture recognition. Moreover, as we will see in the evaluation section, the evaluation dataset and the test procedure used in [4, 5, 16] did not consider gesture variations over the time. Thus their results are likely to be overly optimistic.

Dynamic time warping (DTW) is the core of uWave. It was extensively investigated for speech recognition in the 1970s and early 1980s [8], in particular speaker-dependent speech recognition with a limited vocabulary. Later, HMM-based methods became the mainstream because they are more scalable toward a large vocabulary and can better benefit from a large set of training data. However, DTW is still very effective in coping with limited training data and a small vocabulary, which matches up well with personalized gesture-based interaction with consumer electronics and mobile devices. Wilson and Wilson applied DTW and HMM with XWand [16] to user-independent gesture recognition. The low accuracies, 72% for DTW and 90% for HMM with seven training samples, render them almost impractical. In contrast, uWave focuses on personalized and user-dependent gesture recognition, thus achieving much higher recognition accuracies. It is also important to note that the evaluation data set employed in this work is considerably more extensive than previously reported work, including [4, 5, 16].

It is important to note that some authors use "gesture" to refer to handwritings on touch screen, instead of three-dimensional free-hand movement. Some of these works, e.g. "$1 recognizer" [17], were also based on template matching, similar to uWave. However, because they are based on matching the geometric specifications of two handwritings, it may not apply to matching time series of accelerometer readings, which are subject to temporal dynamics (how fast and forceful the hand moves), three-dimensional acceleration data due to movement of six degrees of freedom, and the confusion introduced by gravity.

III. uWave Algorithm Design

In this section, we present the key technical components of uWave: acceleration quantization, dynamic time warping (DTW), and template adaptation. The premise of uWave is that *human gestures can be characterized by the time series of forces applied to the handheld device*. Therefore, uWave

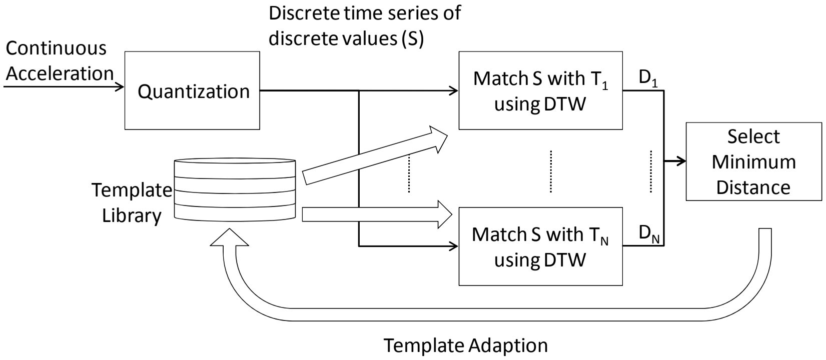


Figure 1: uWave is based on acceleration quantization, template matching with DTW, and template adaptation

bases the recognition on the matching of two time series of forces, measured by a single three-axis accelerometer.

For recognition, uWave leverages a *template library* that stores one or more time series of known identities for every vocabulary gesture, often input by the user. Figure 1 illustrates the recognition process. The input to uWave is a time series of acceleration provided by a three-axis accelerometer. Each time sample is a vector of three elements, corresponding to the acceleration along the three axes. uWave first quantizes acceleration data into a time series of discrete values. The same quantization applies to the templates too. It then employs DTW to match the input time series against the templates of the gesture vocabulary. It recognizes the gesture as the template that provides the best matching. The recognition results, confirmed by the user as correct or incorrect, can be used to adapt the existing templates to accommodate gesture variations over the time.

A. Quantization of Acceleration Data

uWave quantizes the acceleration data before template matching. Quantization reduces the length of input time series for DTW in order to improve computation efficiency. It also converts the accelerometer reading into a discrete value thus reduces floating point computation. Both are desirable for implementation in resource-constrained embedded systems. Quantization improves recognition accuracy by removing variations not intrinsic to the gesture, e.g. accelerometer noise and minor hand tilt.

uWave quantization consists of two steps. In the first step, the time series of acceleration is temporally compressed by an averaging window of 50ms that moves at a 30ms step. This significantly reduces the length of the time series for DTW. The rationale behind it is that intrinsic acceleration produced by hand movement does not change erratically; and rapid changes in acceleration are often caused by noise and minor hand shake/tilt. In the second step, the acceleration data is converted into one of 33 levels, as summarized by Table 1. Non-linear quantization is employed because we find that most samples are between –g and +g and very few go beyond +2g or below –2g.

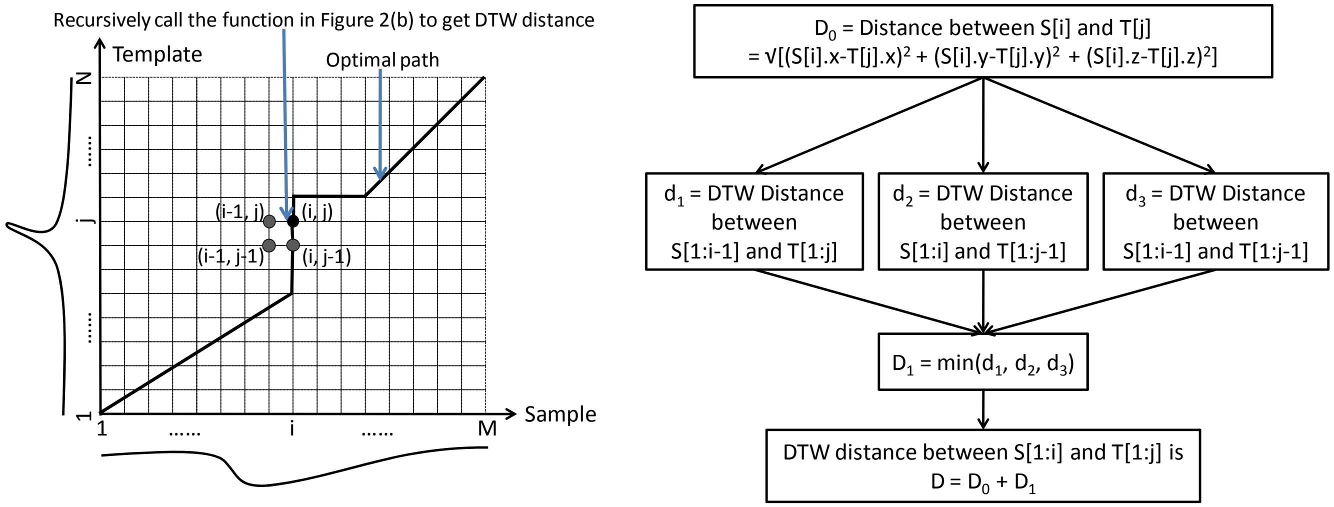
B. Dynamic Time Warping

Dynamic time warping (DTW) is a classical algorithm based on dynamic programming to match two time series with temporal dynamics [8], given the function for calculating the distance between two time samples. uWave employs the Euclidean distance for matching quantized time series of acceleration. Let S[1…M] and T[1…N] denote the two time series. As shown in Figure 2(a), any matching between S and T with time warping can be represented as a monotonic path from the starting point (1, 1) to the end point (M, N) on the M by N grid. A point along the path, say (i, j), indicates that S[i] is matched with T[j]. The matching cost at this point is calculated as the distance between S[i] and T[j]. The path must be monotonic because the matching can only move forward. The similarity between S and T is evaluated by the minimum accumulative distance of all possible paths, or *matching cost*.

DTW employs dynamic programming to calculate the matching cost and find the corresponding optimal path. As illustrated in Figure 2(a), the optimal path from (1, 1) to point (i, j) can be obtained from the optimal paths from (1, 1) to the three predecessor candidates, i.e. (i-1, j), (i, j-1), (i-1, j-1). The matching cost from (1, 1) to (i, j) is therefore the distance at (i, j) plus the smallest matching cost of the predecessor candidates. The algorithm is illustrated in Figure 2(b). The time complexity and space complexity of DTW are both O(M·N).

TABLE 1: UWAVE QUANTIZES ACCELERATION DATA IN A NON-LINEAR FASHION BEFORE TEMPLATE MATCHING

Acceleration Data (a)	Converted Value
a > 2g	16
g < a < 2g	11~15 (five levels linearly)
0 < a < g	1~10 (ten levels linearly)
a = 0	0
-g < a < 0	-1~-10 (ten levels linearly)
-2g < a < -g	-11~-15 (five levels linearly)
a < -2g	-16


Recursively call the function in Figure 2(b) to get DTW distance
Template
Optimal path
N
……
j
……
1
(i-1, j)
(i, j)
(i-1, j-1)
(i, j-1)
1
……
i
……
M
Sample
D$_0$ = Distance between S[i] and T[j]
= √[(S[i].x-T[j].x)^2 + (S[i].y-T[j].y)^2 + (S[i].z-T[j].z)^2]
d$_1$ = DTW Distance between S[1:i-1] and T[1:j]
d$_2$ = DTW Distance between S[1:i] and T[1:j-1]
d$_3$ = DTW Distance between S[1:i-1] and T[1:j-1]
D$_1$ = min(d$_1$, d$_2$, d$_3$)
DTW distance between S[1:i] and T[1:j] is
D = D$_0$ + D$_1$

(a) Graphic illustration of the recursive algorithm

(b) Algorithm for computing the DTW distance between S[1:i] and T[1:j]

Figure 2: Dynamic Time Warping (DTW) algorithm

C. Template Adaptation

As we will show in the evaluation section, there are considerable variations between gesture samples by the same user collected from different days. Ideally, uWave should adapt its templates to accommodate such time variations. Template adaption of DTW for speech recognition has been extensively studied, e.g. [18, 19], and proved to be effective. In this work, however, we only devise two simple schemes to adapt the templates. *Our objective is not to explore the most effective adaptation methods but to demonstrate the template adaptation can be easily implemented and effective in improving recognition accuracy over multiple days.*

Our template adaptation works as follows. uWave keeps two templates generated in two different days for each vocabulary gesture. It matches a gesture input with both templates of each vocabulary gesture and take the smaller matching cost of the two as the matching cost between the input and vocabulary gesture.

Each template has a timestamp of when it is created. On the first day, there is only one training sample, or template, for each gesture. As the user input more gesture samples, uWave updates the templates based on how old the current templates are and how well they match with new inputs. We develop two simple updating schemes. In the first scheme, if both templates for a vocabulary gesture in the library are at least one day old and the input gesture is correctly recognized, the older one will be replaced by the newly correctly recognized input gesture. We refer to this scheme as *Positive Update*. The second scheme differs from the first one only in that we replace the older template with the input gesture when it is incorrectly recognized. We call this scheme *Negative Update*. Positive Update only requires the user to notify uWave when recognition result is incorrect. Negative Update requires the user to point out the correct gesture when a recognition error happens, e.g. by pressing a button corresponding to the identity of the input sample.

IV. Prototype Implementation

We have implemented multiple prototypes of uWave on various platforms, including the Wii remote, Windows Mobile Smartphones, Apple iPhone, and the Rice Orbit sensor [20]. Our accuracy evaluation is based on the Wii remote prototype, due to its popularity and ease of use.

The Wii remote has a built-in three-axis accelerometer from Analog Devices, ADXL330 [21]. The accelerometer has a range of -3g to 3g and noise below 3.5mg when operating at 100Hz [22]. The Wii remote can send the acceleration data and button actions through Bluetooth to a PC in real time. We implement uWave and its variations on a Windows PC using Visual C#. The implementation is about 300 lines of code. The prototype detects the start of a gesture when the ‘A’ button on the Wii remote is pressed; and detects the end when the button is released. While our prototype is based on the Wii remote hardware, uWave can be implemented with any device with a three-axis accelerometer of proper sensitivity and range as are those found in most consumer electronics and mobile devices.

uWave gives out recognition result without perceptible delay in our experiments based on PCs. We measured the speed of uWave implemented in C on multiple platforms. On a Lenovo T60 with 1.6GHz Core 2 Duo, it takes less than 2ms for a template library of eight gestures. On a T-Mobile MDA Pocket PC with Windows Mobile 5.0 and 195MHz TI OMAP processor, it takes about 4ms for the same vocabulary. Such latencies are too short to be perceptible to human users. We also tested uWave on an extremely simple 16-bit microcontroller in the Rice Orbit sensor [20], TI MSP430LF1611. The delay is about 300ms. While this may be perceptible to the user, it is still much shorter than the time a gesture usually takes so that should not impair user experience.

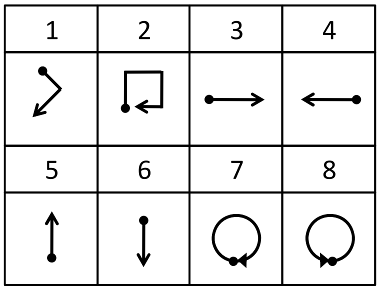


Figure 3: Gesture vocabulary adopted from [6]. The dot denotes the start and the arrow the end

V. EVALUATION

We next present our evaluation of uWave for a vocabulary of predefined gestures based on the Wii remote prototype.

A. *Gesture Vocabulary from Nokia*

We employ a set of eight simple gestures identified by a Nokia research study [4] as preferred by users for interaction with home appliances. The work also provided comprehensive evaluation of HMM-based methods so that a comparison with uWave is possible. Figure 3 shows these gestures as the paths of hand movement.

B. *Gesture Database Collection*

We collect gestures corresponding to the Nokia vocabulary from eight participants with the Wii remote-based prototype. Two of them are undergraduates and others are graduate students; all but one is male. They are in 20s or early 30s, right handed.

The gesture database is collected via the following procedure. For a participant, gestures are collected from seven days within a period of about three weeks. On each day, the participant holds the Wii remote in hand and repeats each of the eight gestures in the Nokia vocabulary ten times. The database consists of 4480 gestures in total and 560 for each participant. This database provides us a statistically significant benchmark for evaluating the recognition accuracy.

It is important to note that the dataset used in [4] consists of 30 samples for each gesture collected from a single user. All of the 30 samples for the same gesture were collected on the same day (the entire dataset of eight gestures were collected over two days). As we will highlight in this work, users exhibit high variations in the same gesture over the time. Samples for the same gesture from the same day cannot capture this and may lead to overly optimistic recognition results.

C. *Recognition without Adaptation*

We first report recognition results for uWave without template adaptation.

1) Test Procedure

Because our focus is personalized gesture recognition, we evaluate uWave using the gestures from each subject separately. That is, the samples from a participant are used to provide templates and test samples for the same subject.

We employ Bootstrapping [23] to further improve the statistical significance of our evaluation. The following procedure applies to each participant separately. For clarity, let us label the samples for each gesture by the order they were collected. For the i^{th} test, we use the i^{th} sample for each gesture from the participant to build eight templates and use the rest samples from the same participant to test uWave. As *i* is from 1 to 70 (10 times by 7 days), we have 70 tests for each participant. Each test produces a confusion matrix that shows the percentage of times how a sample is recognized. We average the confusion matrixes for the 70 tests to produce the confusion matrix for each participant.

We average confusion matrixes of all eight participants to produce the final confusion matrixes. Figure 4 (Left) summarizes the recognition results of uWave over the database for the Nokia gesture vocabulary. In the matrixes, columns are recognized gestures and rows are the actual identities of input gestures.

uWave achieves an average accuracy of 93.5%. Figure 4 (Left) also shows that gesture 1, 2, 6 and 7 have lower recognition accuracy in that they involve similar hand movement as each other, e.g. both gesture 1 and gesture 6 are featured by waving down movement. A closer look into the confusion matrixes for each participant reveals large variation (9%) in recognition accuracy among different participants. *We observed that the participant with the highest accuracy performed the gestures in larger amplitude and slower speed compared to other participants*.

Our evaluation also shows the effectiveness of quantization, i.e., temporal compression and non-linear conversion, of the raw acceleration data. Temporal compression speeds up the recognition by more than nine times without a negative impact on accuracy; and non-linear conversion improves the average accuracy by 1% and further speeds up the recognition.

2) Evaluation using Samples from the Same Day

To highlight how gesture variations from the same user over multiple days impact the gesture recognition, we modify the test procedure above so that when a sample is chosen as the template, uWave is tested only with other samples collected in the same day.

Figure 4 (Right) summarizes the recognition results averaged cross all eight participants. It shows a significantly higher accuracy (98.4%) than that of using samples from all different days. *The difference between Figure 4 (Left) and Figure 4 (Right) highlights the possible variations for the same gesture from the same user over multiple days and the challenge it poses to recognition.* This also indicates that the results reported by some previous work, e.g. [4, 5], were overly optimistic because the evaluation dataset was collected over a very short time.

The same-day accuracy of 98.4% by uWave with one training sample per gesture is comparable to HMM-based methods with 12 training samples (98.6%) reported in [4]. It is worth noting that the accelerometer in Wii remote provides comparable accuracy but larger acceleration range (-3g to 3g) than that used in [4] (-2g to 2g). In reality, however, the

	↙	↲	→	←	↑	↓	↺	↻
↙	92.1	0.1	2.4	1.9	0.1	2.9	0.6	0.1
↲	1.6	91.6	1.3	1.1	0.7	0.4	2.7	0.6
→	0.5	0	95.9	1.2	0.7	1.7	0	0
←	0.3	0	1.6	96.2	0.7	1.1	0	0.1
↑	0.3	0	1.5	0.6	97.0	0.5	0	0.1
↓	2.4	0	2.4	2.3	1.0	91.7	0.1	0
↺	3.4	1.9	2.6	1.7	0.4	0.7	89.2	0
↻	1.1	0.6	1.7	0.9	0.8	0.7	0	94.2

	↙	↲	→	←	↑	↓	↺	↻
↙	98.4	0	0.3	0.4	0	0.4	0.3	0.2
↲	0.5	98.3	0.2	0	0.3	0.1	0.4	0.1
→	0.2	0	98.3	0.6	0.1	0.6	0.2	0
←	0.2	0	0.3	98.8	0.3	0.2	0.2	0
↑	0.4	0	0.2	0.4	98.7	0.1	0.2	0
↓	0.7	0	0.6	0.5	0.3	97.7	0.2	0
↺	0.5	0.4	0.4	0.1	0.1	0.3	98.1	0.2
↻	0.2	0.1	0.1	0.2	0	0	0.2	99.2

Figure 4: Confusion matrixes for the Nokia vocabulary without adaptation. Columns are recognized gestures and rows are the actual identities of input gestures. (Left) Tested with samples from all days (average accuracy is 93.5%); (Right) Tested with samples from the same day as the template (average accuracy is 98.4%)

	↙	↲	→	←	↑	↓	↺	↻
↙	96.8	0	1.5	0.3	0	1.1	0	0.2
↲	0.7	96.4	0.5	0.2	0.2	0.4	1.2	0.5
→	0	0	98.9	0.6	0	0.5	0	0
←	0.2	0	0.3	98.9	0.2	0.5	0	0
↑	0.2	0	0.2	0.1	99.3	0.2	0	0
↓	0.6	0	0.6	0.3	1.7	96.8	0	0
↺	0.8	2.0	2.0	0.4	0	0.2	94.6	0
↻	1.0	0.4	1.1	0.4	0	0	0	97.1

	↙	↲	→	←	↑	↓	↺	↻
↙	97.7	0	1.2	0.6	0	0.6	0	0
↲	0.6	98.6	0.2	0.1	0	0.1	0.3	0.1
→	0.1	0	99.1	0.4	0.1	0.4	0	0
←	0.1	0	0.4	99.0	0.1	0.4	0	0
↑	0.2	0	0.3	0.1	99.2	0.2	0	0
↓	0.5	0	0.4	0.2	0.5	98.3	0	0.1
↺	0.4	0.5	07	0.2	0.1	0.2	98.0	0
↻	0.2	0	0.3	0.4	0.1	0.1	0	98.9

Figure 5: Confusion matrixes for the Nokia vocabulary with adaptation, tested with samples from all days. Columns are recognized gestures and rows are the actual identities of input gestures. (Left) Positive Update (average accuracy is 97.4%); (Right) Negative Update (average accuracy is 98.6%)

acceleration produced by hand movement rarely exceeds the range from -2g to 2g. Hence, the impact of difference in the accelerometers on the accuracy should be insignificant.

D. Recognition with Adaptation

The considerable difference between Figure 4 (Left) and Figure 4 (Right) motivates the use of template adaptation to accommodate variations over the time in order to achieve accuracy close to that in Figure 4 (Right). We report the results next.

Again, we evaluate uWave with adaptation for each participant separately. Because the adaption is time-sensitive, we have to apply Bootstrapping in a more limited fashion. Let us label the days in which a participants' gestures were collected by the time order, from one to seven. For the *i*th test, we assume the evaluation starts on the *i*th day and applies the template adaptation in the following days, from (*i+1*)th to 7th and then from 1st to (*i-1*)th. We have seven tests for each participants and each produces a confusion matrix. We average them to produce the confusion matrix for each participant and average the confusion matrixes of all participants for the final one.

Figure 5 summarizes the recognition results averaged across all participants. It shows an accuracy of 97.4% for Positive Update and 98.6% for Negative Update, significantly higher than that without adaptation (Figure 4 Left) and close to that tested with samples from the same day (Figure 4 Right). While template adaptation requires user feedback when a recognition error happens, the high accuracy indicates that it is needed only for 2-3% of all the test samples.

VI. uWave-Enhanced Applications

In this section, we present two applications that have been enhanced with the uWave technology, one for gesture-based

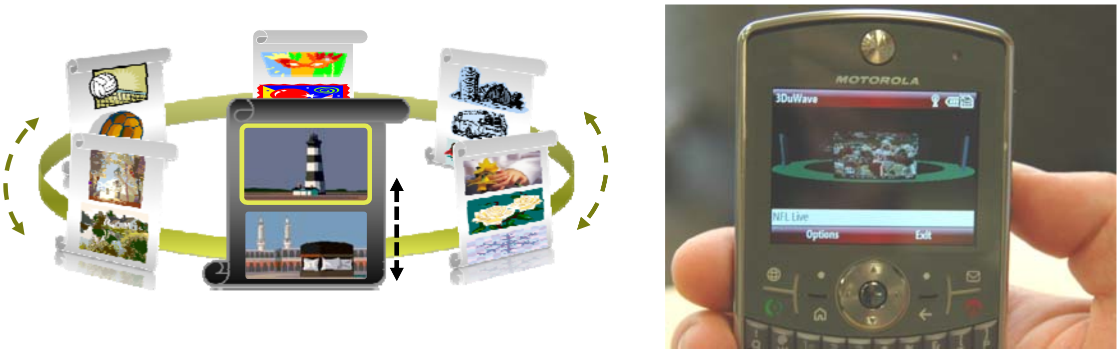


Figure 6: Mobile 3D User Interface with uWave-based gesture interaction: (left) Illustration of the user interface and (right) prototype implementation

user authentication and the other for gesture-based interaction with mobile phones.

A. Gesture-based Light-Weight User Authentication

Personalization is a growing component of many multi-user systems these days. However, outside the traditional realm of password-based strong authentication, there is a need for light-weight authentication techniques that prioritize ease-of-use over hard security. Under many scenarios, user-specific data can be privacy insensitive. For example, personal profiles or personalized configurations on a TV remote shared by family members are likely to be privacy-insensitive. For privacy-insensitive user-specific data, this manner of light-weight, 'soft' user authentication provides a mechanism for a user to personalize the device. The objectives are 1) accurate recognition of a user and 2) to be user-friendly, easy to remember and easy to perform. While many paradigms exist for user authentication, including password-based and biometrics, uWave enables authentication based on physical manipulation of the device with low cost and high efficiency. It is particularly suitable for implementation on resource-constrained devices, such as mobile phones and TV remotes.

We conducted two user studies with the Wii remote-based prototype described in Section 4. We assume a small number (<10) of users share a device that can be personalized by loading user-specific data. The target audience for these devices is primarily shared consumer electronics, as mentioned above. The participants chose their signature collectively in the first user study and independently in the second. Each study involved five participants who are Rice undergraduate and graduate students.

These studies show that uWave can recognize user-defined gestures with higher than 99.5% accuracy in both selection procedures. In the follow-up survey, the participants rated the difficulty of memorizing the gestures as 1.4 on average (on a 1-5 scale where 1 is "easiest" and 5 is "most difficult"), compared to 2.2 for memorizing a user ID (smaller number means less difficult); they rated the difficulty of performing the gesture as 1.9 on average, about the same as that for typing in a user ID. These results show that the usability of gesture-based schemes is at least as good as traditional user ID/password-based ones for authentication in terms of their cognitive and kinetic load on the user.

It is important to note that the authors of [24-26] investigated gestures as a biometrics for 'hard' user authentication where security is important. They attempted to recognize the user based on how she performs a given gesture. In contrast, our application of uWave is targeted at 'soft' user authentication with any user-defined gestures.

B. Gesture-based 3D Mobile User Interface

One of the strengths of uWave is that it can recognize three-dimensional hand movement. It has been shown that it is intuitive and convenient to navigate a 3D user interface with 3D hand gestures [27]. Qualitatively, being able to manipulate a 3D interface using a 3D gesture is much more compelling than traditional button-based solutions. In order to explore this, we developed a 3D-mobile application and integrated uWave with it to enable gesture-based navigation.

The 3D application was built around a social networking-based video-sharing service under development within Motorola. The interface shows a rotating ring that contains thumbnails of various users (a friends list) as in Figure 6. Additionally, upon selecting a particular user, one can scroll through different video clips that have been submitted by that user. We employed uWave to navigate this user interface using a series of specific movements such as tilting and slight shaking, which are more appropriate for a mobile device when the user is focused on the screen. We also added the personalization features of uWave to allow users to re-map gestures to their liking, enabling custom navigation of the 3D interface.

The application runs on an accelerometer-enhanced Smartphone, and is implemented in C++ for the Windows Mobile 6 Platform. The 3D interface is built and rendered using the Mobile 3D Graphics (M3G) API. The acceleration data is read via Bluetooth Serial at 100Hz. Even when the 3D rendering consumes a significant amount of memory, uWave works smoothly with it, without introducing any human perceptible performance degradation.

VII. Discussion

We next address the limitations of uWave and gesture recognition based on accelerometers in general.

A. Gestures and Time Series of Forces

Due to a lack of a standardized gesture vocabulary, human users may have diverse opinions on what constitutes a unique gesture. As noted early, the premise of uWave is that human gestures can be characterized as time series of forces applied to a handheld device. Therefore, the temporal dynamic of gestures is closer to speech than to handwriting, which is usually recognized as the final contours without regard to the time sequence of the contours. However, it is important to note that while one may produce the three-dimensional contour of the hand movement given a time series of forces, the same contour may be produced by very different time series of forces. Nevertheless, our evaluation gesture samples were collected without enforcing any definition of gestures to our participants. The high accuracy of uWave indicates that its premise is close to how users perceive gestures and how users perform gestures.

B. Challenge of Tilt

On the other hand, uWave relies on a single three-axis accelerometer to infer the force applied. However, *the reading of the accelerometer does not directly reflect the external force*, because the accelerometer can be tilted around three axes. The same external force may produce different accelerations along the three axes of the accelerometer if it is tilted differently; likewise, the different forces may also produce the same accelerometer readings. Only if the tilt is known, the force can be inferred from the accelerometer readings.

The opportunity for detecting the tilt during hand movement is very limited with a single accelerometer. We attempted to address it by allowing each pair of matching points on the DTW grid (See Figure 2) to calculate the distance based on tilts of small angles. While it helped with matching samples of the same gesture collected with different tilts, it also increased the confusion between certain gestures, largely due to the confusion between gravity and the external force. To fully address tilt variation, extra sensors, e.g. compass and gyroscope, will be necessary for additional information.

C. User-Dependent vs. User Independent Recognition

This work and numerous others are targeted at user-dependent gesture recognition only. The reasons are multiple. First, user-independent gesture recognition is difficult. Our database shows great variations among participants even for the same predefined gesture. For example, if we treat all the samples in the database as from the same participant and repeat our bootstrapping test procedure, the accuracy will decrease to 75.4% from 98.4% for user-dependent recognition. To improve the accuracy of user-independent recognition, a large set of training samples and a statistical method are necessary. More importantly, research is required to identify the common "features" from the acceleration data for the same gesture. In speech recognition, MFCC and LPCC have been found to capture the identity of speech very effectively. Unfortunately, we do not know their counterparts for acceleration-based gesture recognition. Second, user-independent gesture recognition may not be as attractive as speaker-independent speech recognition because there is no standard or commonly accepted gestures for interaction. Commonly recognized gestures by humans are often simple, such as those in the Nokia vocabulary. As they are short and simple, however, they can be easily confused with each other, in particular with the presence of tilt and user variations. On the other hand, for personalized gestures composed by users, it is almost impossible to collect a large dataset for statistical methods to be effective.

D. Gesture Vocabulary Selection

The confusion matrixes presented in Figure 4 and Figure 5 highlight the importance of selecting the right gesture vocabulary for higher accuracy. As from Figure 4, we can see that uWave often confuses Gesture 1 with Gesture 7. The reason is that tilt of the handheld device can transform different forces into similar accelerometer readings. Unlike speech recognition, gesture recognition has more flexible inputs, because the user can compose gestures without the constraint of a "language". More complicated gestures may lead to higher accuracy because they are likely to have more features that distinguish them from each other, in particular, offsetting the effect of tilt and gravity. Nevertheless, complicated gestures pose a burden to human users: the user has to remember how to perform complicated gestures in a consistent manner and associate them with some unrelated functionality. Eventually, the number of complicated gestures a user can comfortably command may be quite small. This may limit gesture-based interaction with a relatively small vocabulary for which uWave indeed excels.

VIII. CONCLUSIONS

We present uWave for interaction based on personalized gestures and physical manipulations of a consumer electronic or mobile device. uWave employs a single accelerometer so it can be readily implemented on many commercially available consumer electronics and mobile devices. The core of uWave includes dynamic time warping (DTW) to measure similarities between two time series of accelerometer readings; quantization for reducing computation load and suppressing noise and non-intrinsic variations in gesture performance; and template adaptation for coping with gesture variation over the time. Its simplicity and efficiency allow implementation on a wide range of devices, including simple 16-bit microcontrollers.

We evaluate the application of uWave to user-dependent recognition of predefined gestures with over 4000 samples collected from eight users over multiple weeks. Our experiments demonstrate that uWave achieves 98.6% accuracy starting with only one training sample. This is comparable to the reported accuracy by HMM-based methods [4] with 12 training samples (98.9%). We show that the quantization improves recognition accuracy and reduces the computation load. Our evaluation also highlights the challenge of variations over the time to user-dependent gesture recognition and the challenge of variations across users to user-independent gesture recognition. We presented two applications of uWave: gesture-based authentication and mobile 3D interface with gesture-based navigation on an accelerometer-enhanced Smartphone. Both applications show high recognition accuracy and recognition speed with different hardware features and system resources.

We believe uWave is a major step toward building technology that facilitates personalized gesture recognition. Its accurate recognition with one training sample is critical to the adoption of personalized gesture recognition in a range of devices and platforms and to the realization of novel gesture-based navigation of next generation user interfaces.

Acknowledgments

The work is supported in part by NSF awards CNS/CSR-EHS 0720825 and IIS/HCC 0713249 and by a gift from Motorola Labs. The authors would like to thank the participants in our user studies and anonymous reviewers whose comments helped improve the final version.

References

[1] T. Baudel and B.-L. Michel, "Charade: remote control of objects using free-hand gestures," *Commun. ACM*, vol. 36, pp. 28-35, 1993.

[2] X. Cao and R. Balakrishnan, "VisionWand: interaction techniques for large displays using a passive wand tracked in 3D," in *Proc. ACM Symp. User Interface Software and Technology (UIST)*. Vancouver, Canada: ACM, 2003.

[3] J. K. Perng, B. Fisher, S. Hollar, and K. S. J. Pister, "Acceleration sensing glove (ASG)," in *Digest of Papers for Int. Symp. Wearable Computers*, 1999, pp. 178-180.

[4] J. Kela, P. Korpipää, J. Mäntyjärvi, S. Kallio, G. Savino, L. Jozzo, and D. Marca, "Accelerometer-based gesture control for a design environment," *Personal Ubiquitous Computing*, vol. 10, pp. 285-299, 2006.

[5] J. Mäntyjärvi, J. Kela, P. Korpipää, and S. Kallio, "Enabling fast and effortless customisation in accelerometer based gesture interaction," in *Proc. Int. Conf. Mobile and Ubiquitous Multimedia*. College Park, MA: ACM, 2004.

[6] L. R. Rabiner and B. H. Juang, "An Introduction to Hidden Markov Models," in *IEEE ASSP Magazine*, 1986, pp. 4-15.

[7] Y. Wu and T. S. Huang, "Vision-Based Gesture Recognition: A Review," in *Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction*: Springer-Verlag, 1999.

[8] C. S. Myers and L. R. Rabiner, "A comparative study of several dynamic time-warping algorithms for connected word recognition," *The Bell System Technical Journal*, vol. 60, pp. 1389-1409, 1981.

[9] Nintendo, "Nintendo Wii," in *http://www.nintendo.com/wii/*.

[10] J. Liu, Z. Wang, L. Zhong, J. Wickramasuriya, and V. Vasudevan, "Demonstration: uWave: Accelerometer-based personalized gesture recognition," in *ACM Symp. User Interface Software and Technology (UIST)*, 2008.

[11] G. Heumer, H. B. Amor, M. Weber, and B. Jung, "Grasp recognition with uncalibrated data gloves - A comparison of classification methods," in *IEEE Virtual Reality Conference*, 2007, pp. 19.

[12] S. Ronkainen, J. Häkkilä, S. Kaleva, A. Colley, and J. Linjama, "Tap input as an embedded interaction method for mobile devices," in *Proc. Int. Conf. Tangible and Embedded Interaction*. Baton Rouge, LA: ACM, 2007.

[13] I. J. Jang and W. B. Park, "Signal processing of the accelerometer for gesture awareness on handheld devices," in *Proc. IEEE Int. Wkshp. Robot and Human Interactive Communication*, W. B. Park, Ed., 2003, pp. 139-144.

[14] P. Keir, J. Payne, J. Elgoyhen, M. Horner, M. Naef, and P. Anderson, "Gesture-recognition with non-referenced tracking," in *IEEE Symp. 3D User Interfaces (3DUI)*, 2006, pp. 151.

[15] AiLive Inc, "AiLive LiveMove Pro," *http://www.ailive.net/liveMovePro.html*.

[16] D. Wilson and A. Wilson, "Gesture recognition using XWand," Robotics Institute, Carnegie Mellon University 2004.

[17] J. O. Wobbrock, A. D. Wilson, and Y. Li, "Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes," in *Proc. ACM Symp. User Interface Software and Technology (UIST)*, 2007.

[18] F. R. McInnes, M. A. Jack, and J. Laver, "Template adaptation in an isolated word-recognition system," *IEE Proceedings*, vol. 136, 1989.

[19] R. Zelinski and F. Class, "A learning procedure for speaker-dependent word recognition systems based on sequential processing of input tokens," in *Proc. IEEE ICASSP*, 1983.

[20] Rice Efficient Computing Group, "Rice Orbit Sensor Platform," in *http://www.recg.org/orbit.htm*.

[21] H. Wisniowski, "AnalogDevices and Nintendo collaboration drives video game innovation with iMEMS motion signal processing technology," *Analog Devices*, 2006.

[22] Analog Device, "Small, low power, 3-Axis ±3g i MEMS® accelerometer: ADXL330 datasheet," 2006.

[23] M. R. Chernick, *Bootstrap: A practitioner's guide.*, 1999.

[24] E. Farella, S. O'Modhrain, L. Benini, and B. Riccó, "Gesture Signature for Ambient Intelligence Applications: A Feasibility Study," in *Pervasive Computing*, 2006, pp. 288-304.

[25] K. Matsuo, F. Okumura, M. Hashimoto, S. Sakazawa, and Y. Hatori, "Arm Swing Identification Method with Template Update for Long Term Stability," in *Advances in Biometrics*, 2007, pp. 211-221.

[26] F. Okumura, A. Kubota, Y. Hatori, K. Matsuo, M. Hashimoto, and A. Koike, "A Study on Biometric Authentication based on Arm Sweep Action with Acceleration Sensor," in *Int. Symp. Intelligent Signal Processing and Communications*, 2006, pp. 219-222.

[27] K. Hinckley, J. Tullio, R. Pausch, D. Proffitt, and N. Kassell, "Usability analysis of 3D rotation techniques.," in *Proc. ACM Symp. User Interface Software and Technology (UIST)*, 1997.