

uWave: Accelerometer-based Personalized Gesture Recognition and Its Applications

Vijay Sukhadeve
Computer Science Dept.
Worcester Polytechnic Institute (WPI)

INTRODUCTION

- This paper-work presents an opportunity for spontaneous interaction with consumer electronics and mobile devices based on gestures or physical manipulation of the devices.
- uWave is presented to address the multiple technical challenges faced by gesture-based interaction and focus on gestures without regard to finger movement such as sign language.
- Unlike statistical methods, uWave requires a single training sample for each gesture pattern and it 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 delivers competitive accuracy.
- 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).
- Effective implementation with multiple prototypes of uWave on various platforms, including Smartphones, microcontroller, and the Nintendo Wii remote hardware.

RELATED WORK

- the Wii remote has a “camera” (IR sensor) inside the remote which detects motion by tracking the relative movement of IR transmitters mounted on the display. It basically translates a “gesture” into “handwriting”, contributing itself to a rich set of handwriting recognition techniques.
- “smart glove” based solutions 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. Therefore, they are unfit for spontaneous interaction due to the high overhead of engagement.
- LiveMove Pro 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.
- Dynamic time warping (DTW) is the core of uWave. Extensively investigated for speech recognition, in particular speaker- dependent speech recognition with a limited vocabulary. 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.

UWAVE ALGORITHM DESIGN

- The key technical components of uWave: acceleration quantization, dynamic time warping(DTW), and template adaptation.
- uWave 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.
- 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 & then templates into a time series of discrete values. 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.

Quantization of Acceleration Data

- 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.
- Quantization improves recognition accuracy by removing variations not intrinsic to the gesture, e.g. accelerometer noise and minor hand tilt.

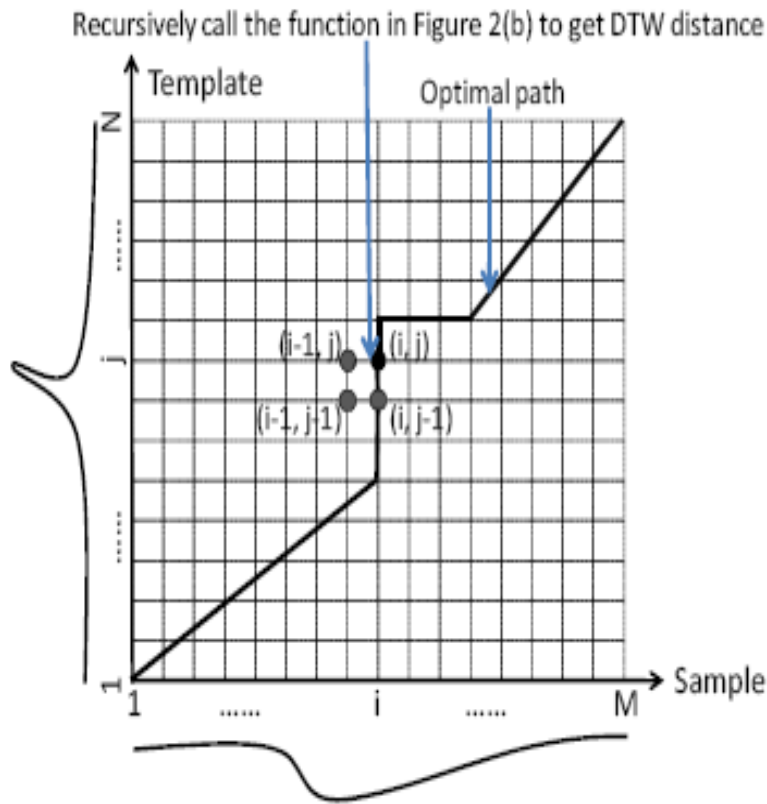
TABLE 1: UWAVE QUANTIZES ACCELERATION DATA IN A NON-LINEAR FASHION BEFORE TEMPLATE MATCHING

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

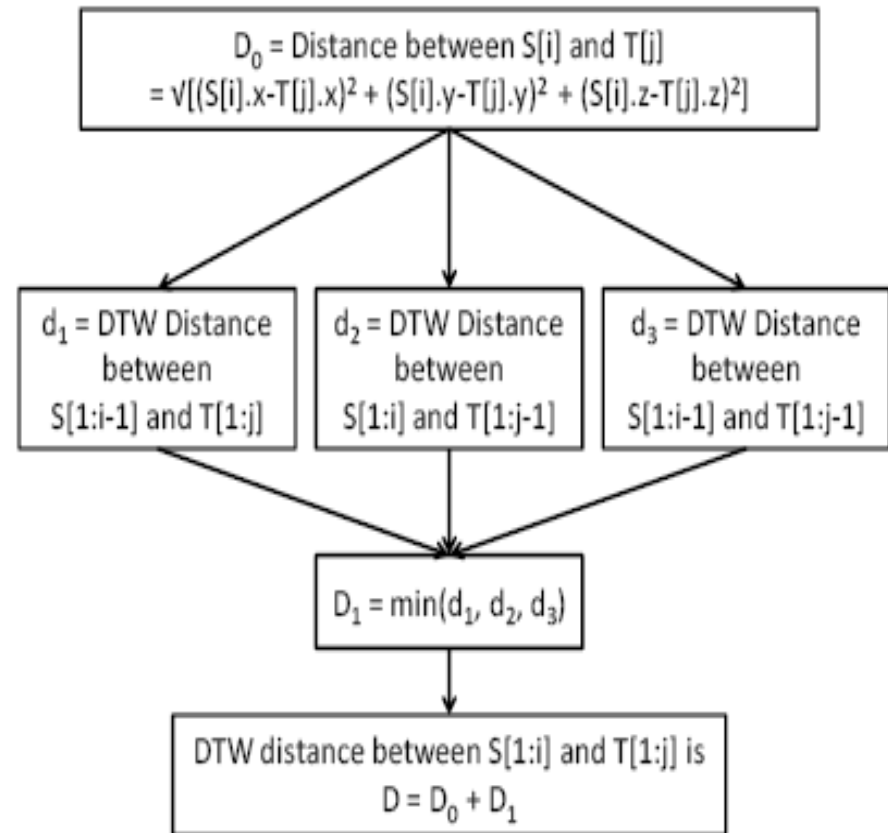
Dynamic Time Warping

- Dynamic time warping (DTW) is a classical algorithm based on dynamic programming to match two time series with temporal dynamics , given the function for calculating the distance between two time samples.
- uWave employs the Euclidean distance for matching quantized time series of acceleration.
- DTW employs dynamic programming to calculate the matching cost and find the corresponding optimal path.

Dynamic Time Warping (DTW) algorithm



(a) Graphic illustration of the recursive algorithm



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

Template Adaptation

- 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 takes 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. 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.

PROTOTYPE IMPLEMENTATION

- multiple prototypes of uWave have been implemented on various platforms, including the Wii remote, Windows Mobile Smartphones, Apple iPhone, and the Rice Orbit sensor.
- 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 . The accelerometer has a range of -3g to 3g and noise below 3.5mg when operating at 100Hz . The Wii remote can send the acceleration data and button actions through Bluetooth to a PC in real time.

EVALUATION

- Evaluation of uWave for a vocabulary of predefined gestures based on the Wii remote prototype -----

A. Gesture Vocabulary from Nokia

- Team employed a set of eight simple gestures identified by a Nokia research study 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 C** shows these gestures as the paths of hand movement.

B. Gesture Database Collection

- gestures are collected corresponding to the Nokia vocabulary from eight participants with the Wii remote-based prototype. 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. 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*** (Refer **Figure A**)

- *1) Test Procedure*

- Because focus is on personalized gesture recognition, uWave is evaluated using the gestures from each subject separately. 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.
- A closer look into the confusion matrixes for each participant reveals large variation (9%) in recognition accuracy among different participants.
- The evaluation results also show 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, uWave is tested only with other samples collected in the same day.
- Figure (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 (Left) and Figure (Right) highlights the possible variations for the same gesture from the same*

- ***D. Recognition with Adaptation*** (Refer Figure B)
- The considerable difference between Figure A (Left) and Figure A (Right) motivates the use of template adaptation to accommodate variations over the time in order to achieve accuracy close to that in Figure A (Right).
- uWave is evaluated with adaptation for each participant separately. Because the adaption is time-sensitive, we have to apply Bootstrapping in a more limited fashion.
- 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 B 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 A Left) and close to that tested with samples from the same day (Figure A 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.

UWAVE-ENHANCED APPLICATIONS

- ***A. Gesture-based Light-Weight User Authentication***

- 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.

- ***B. Gesture-based 3D Mobile User Interface***

- uWave can recognize three dimensional hand movement and it has been observed that it is intuitive and convenient to navigate a 3D user interface with 3D hand gestures .
- Manipulating a 3D interface using a 3D gesture is much more compelling than traditional button-based solutions. In order to explore this, team developed a 3D-mobile application and integrated uWave with it to enable gesture-based navigation.

















DISCUSSION

- The limitations of uWave and gesture recognition based on accelerometers in general -----
- ***A. Gestures and Time Series of Forces***
- The premise of uWave is that human gestures can be characterized as time series of forces applied to a handheld device. 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.
- ***B. Challenge of Tilt***
- 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.

C. User-Dependent vs. User Independent Recognition

- First, user-independent gesture recognition is difficult. The database shows great variations among participants even for the same predefined gesture. 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 reason for confusion is that tilt of the handheld device can transform different forces into similar accelerometer readings. 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.

Figure A 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%)

								
	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 B 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%)

								
	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





















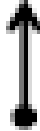
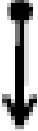


								
	97.7	0	1.2	0.6	0	0.6	0	0
	0.6	98.6	0.2	0.1	0	0.1	0.3	0.1
	0.1	0	99.1	0.4	0.1	0.4	0	0
	0.1	0	0.4	99.0	0.1	0.4	0	0
	0.2	0	0.3	0.1	99.2	0.2	0	0
	0.5	0	0.4	0.2	0.5	98.3	0	0.1
	0.4	0.5	0.7	0.2	0.1	0.2	98.0	0
	0.2	0	0.3	0.4	0.1	0.1	0	98.9

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

1	2	3	4
			
5	6	7	8
			

CONCLUSIONS

- uWave presented for interaction based on personalized gestures and physical manipulations 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.
- Evaluation of uWave using a large gesture library with over 4000 samples collected from eight users over multiple weeks 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.
- 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.

- Two applications of uWave are presented: 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.
- From the perspective of developing technology , uWave is a major step and holds out good prospects 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.

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*,

- [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 $\pm 3g$ 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. Ricc , "Gesture Signature for Ambient Intelligence Applications: A Feasibility Study," in *Pervasive Computing*, 2006, pp. 288-