Capturing Basic Movements for Mobile Gaming Platforms Embedded with Motion Sensors ☆

A. Kailas[a,*], C. -C. Chong[b], F. Watanabe[c]

[a]*Dept. of Electrical and Computer Engineering, University of North Carolina Charlotte, 9201 University City Bvd., Charlotte, NC, 28223-0001*
[b]*France Telecom R&D San Francisco, LLC, 801 Gateway Blvd, Suite 500, South San Francisco, CA 94080*
[c]*DOCOMO Communications Laboratories USA, Inc., 3240 Hillview Avenue, Palo Alto, CA 94304-1201*

Abstract

A novel experimental setup using accelerometer and gyroscope sensors embedded on a single board along with a distance-based pattern recognition algorithm is presented for accurately identifying basic movements for possible application in gaming using a mobile platform. As an example, we considered some basic step sequences in the popular dance game (e.g., dance dance revolution), and could detect these movements with a reasonably high probability. We envision that the experimental results presented in this paper will motivate future research in the world of mobile gaming applications using advanced smart phones with a dual module design.

Keywords: Dance dance revolution, human pose recognition, mobile gaming, mobile sensors

1. Introduction: Evolution of Mobile Gaming

Simply stated, a mobile game is a video game that is played on a mobile phone using the technologies present on the device itself. This classification does not include handheld video systems/products such as PlayStation Portable or Nintendo DS. Mobile games are

☆This work was done when Aravind Kailas was a Visiting Research Scientist at DOCOMO USA Labs, and when Chia-Chin Chong was a Senior Research Engineer at DOCOMO USA Labs.

*Corresponding author

Email addresses: aravindk@ieee.org (A. Kailas), chiachin.chong@orange-ftgroup.com (C. -C. Chong), watanabe@docomolabs-usa.com (F. Watanabe)

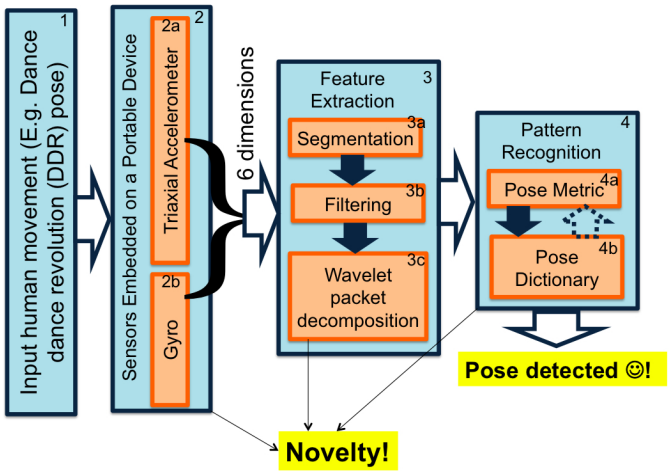


Figure 1: Block diagram highlighting the main contributions of the paper.

applications that simply use the device platform to run the game software. These games maybe downloaded on the fly and installed while mobile, or onto the handset using a cable, or come pre-installed by the original equipment manufacturer (OEM) or by the cellular operator as applications or services. Mobile games are a growing market [1] and according to [2], in 2010 the sales of smartphones has been booming. More than 10 million people worldwide play games on mobile phones and handheld devices [3], and the world-wide mobile gaming revenue is expected to reach $11.4 billion by 2014. These are very important motivations for game developers and designers to create blockbuster games for mobile platforms.

The contributions of the paper can be summarized as follows:

1. Experimental Setup: A novel experimental setup using accelerometer and gyroscope sensors embedded on a single board (Block 2 in Fig. 1);
2. Data analysis and Feature Extraction: Our feature extraction using wavelet packet decomposition (WPD), thereby providing better resolution for analysis than the fast Fourier transform (FFT) (Block 3c in Fig. 1);
3. Algorithm Development: Pattern recognition-based algorithm based on the Mahalanobis distance, which is based on the correlations between the data points (Block 4a in Fig. 1);
4. Pose Recognition: Reliable pose recognition using two sensor boards (i.e., accelerometer

and gyro combination) for games with a lot of human movements (e.g., dance dance revolution (DDR), a very popular game), is demonstrated with a high probability of success (Block 4 in Fig. 1) to motivate future research game changing phones with a dual module design such as the Fujitsu F-04B.

The rest of the paper is organized as follows. Section II described the prior works. Section III outlines the methodology and experiments conducted. Experimental results are discussed in Section IV and conclusions are presented in Section V.

2. Prior Works

Human motion capture consists of the recording of human body movements for immediate or delayed analysis and playback [4]. Three axis accelerometers have been used to (1) monitor frail activity [5]; (2) estimate energy expenditure [6]; (3) track articulated human motion [7]; (4) detect knee unlock [8]; (5) detect gesture awareness [9]; (6) analyze human motion [10]; and (7) measure heart motion [11], to list a few. Inertial measurement unit (IMU) modules comprising accelerometer and angular rate sensors have been used to distinguish both low- and high-speed human activities [12]. Combinations of accelerometers and gyroscope sensors have been used to (1) estimate human body orientation [13]; (2) detect human posture and walking speed [14], and (3) human recognition via gait recognition [15], to name a few. However, none of these implementations are for a portable device such as a mobile phone. Stated simply, the users need to wear some sensors on specific parts of their bodies [16], and the feature extraction is suitable only for a limited range of user context.

There has been some work on developing gait analysis and posture recognition algorithms for a mobile platform [17] (and the references within). For instance, Iso and Yamazaki demonstrated gait analysis on a cell phone with a single three-axis accelerometer [18], [19]. Karantonis *et. al.*, presented an implementation of a real-time classification system for the types of human movement using a single triaxial accelerometer module worn at the hip, and this system distinguished between activity and rest, measured the tilt angle of the body, and could detect walking and falling conditions [20]. Recently, a multi-person pose recognition

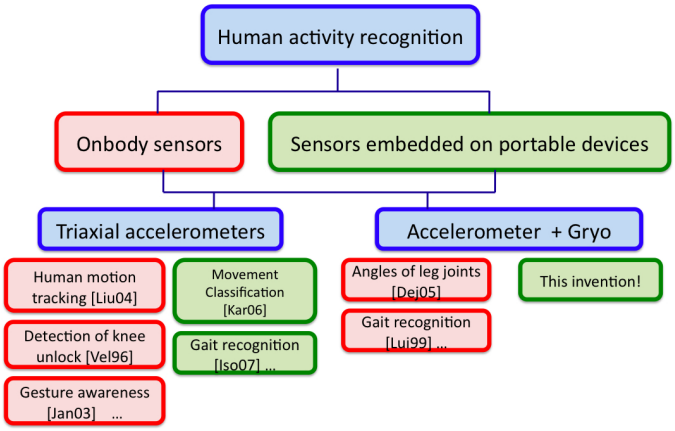


Figure 2: Classification of the prior works in human activity and pose recognition.

module comprising a triaxial accelerometer, a microcontroller, and a ZigBee chip for mobile robots was implemented by Song and Chen [21]. However, to the best of our knowledge, there have been no works that consider a combination of accelerometer and gyroscopes on a mobile device. An illustration of this classification and differences between the prior works and this paper is shown in Fig. 2.

To the best of our knowledge, there has been only one work that deploys accelerometers for video games such as the dance dance revolution (DDR). Crampton *et. al.*, created and tested a wearable sensor network that detects the subjects body position as input for the video games [22]. The effects of multiple accelerometers on accurate pose detection support the use of accelerometer-based sensor networks as input in video games. This paper is different because, a combination of accelerometer and gyro sensors embedded on a portable dual mode mobile platform is considered for pose recognition resulting motivating future research in mobile gaming. Moreover, reasonable pose recognition is demonstrated using fewer sensors and our classification algorithm.

3. Methodology and Experimental Setup

3.1. Accelerometer-Gyro Board (Experimental Device)

The device we used for collecting data is shown in Fig. 3. The device integrates an IHU-500 board from InvenSense [23], an SD card with 2 GB capacity, and a USB port.

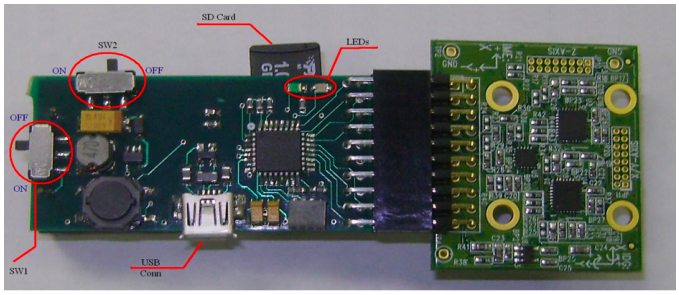


Figure 3: The integrated sensor board mobile platform.

The IHU-500 board is equipped with an integrated triaxis accelerometer (IME-3000), a two-dimension gyroscope (IDG-500), and a one-dimension gyroscope (ISZ-500). Therefore, the device can capture acceleration and rotation in three dimensions. In the futureThe sampling rate is set to be 500 Hz. Up to 8-hour sampling data with 16 bits/sample/dimension can be collected and stored in the in the 2 GB SD card. Up to 60 data sets could be recorded after which the data must be uploaded to a PC. Combined duration of the 60 datasets did not exceed eight hours of data recordings.

3.2. Pose Recognition Experiments

Three different experiments were conducted for pose recognition by using the DDR as an example, which depends on the placement of the accelerometer gyro on the body. It is vital to get the position and rotation of the "points of motion" to infer about the DDR step. They are explained as follows:

Experiment 1: Every DDR song sequence can be thought of as a combination of distinct sequences, each being a combination of forward, backward, right, and left movements. With this in mind, Experiment 1 entailed executing 48 poses, each pose being a permutation of the four distinct movements without any repetition. For Experiment 1, 100 readings are taken each time with a one-two second period between each movement. For this experiment, one sensor board was strapped tightly to the ankle, while the other was strapped around the torso (or the waist).

Experiment 2: In the experiment, one sensor board was strapped tightly to the ankle, while the other was positioned on the palm. A total of five poses were executed, and they

were right, forward, backward, forward-backward, and backward-forward. The objective of this experiment was to distinguish the short movements or pose changes.

Experiment 3: In the experiment, one sensor board was strapped tightly to the ankle, while the other was strapped around the torso (or the waist). A total of five poses were executed, and they were right, forward, backward, forward-backward, and backward-forward. Like Experiment 2, the objective here also was to distinguish the short movements or pose changes.

3.3. Training Phase

The purpose of the training phase is to create a raw data set of a given number of poses for an individual. In order to record a pose, after sufficient practice, the subject performs each pose 50 times. A reading consisted of the x, y, and z accelerations, and the pitch, roll, and yaw from the gyroscope from both the sensor boards. A data set for training comprised raw data from a single subject. In order to convert a set of raw data into recognizable pose pattern, two basic calculations must be performed. The mean acceleration for each pose must be computed. The acceleration values and standard deviations for each axis of the accelerometer for each pose are stored in a pose bank for subsequent tasks. This is repeated for every subject. The training phase and developing a customized pose dictionary are essential for pose classification. A high-level flowchart illustration the use of training data to generate a generic pose dictionary for the application is shown in Fig. 4.

3.4. Feature Extraction Phase

The proposed feature extraction framework has the following steps:

1. *Segmentation*: This step is used to determine when the gesture begins and when it ends. Segmentation is used mainly to automatically determine the beginning and end of the pose. Based on the experiments and observation of the data from the accelerometers, a pose is defined as follows. A pose begins with a fast acceleration, a continuous change in the direction throughout the duration of the pose, and it ends with a stop of the movement. One approach is to do this manually, i.e., every new set of data is preceded

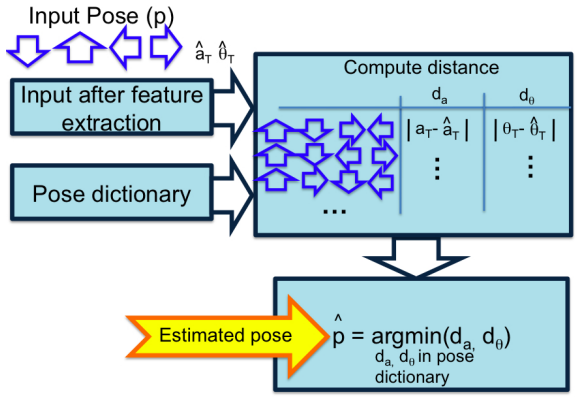


Figure 4: High-level flowchart of the classification algorithm.

and followed by a sudden jerk. With sufficient data points, an empirical approach was devised and is described below. In order to correctly segment the pose based on accelerometer data, one needs to check the magnitude of the parameter, $\mathcal{D}$, defined as

$$\mathcal{D} = \sqrt{(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2 + (z_k - z_{k-1})^2}, \tag{1}$$

where x_k, y_k, and z_k are the values of the accelerometer at instant k. If $\mathcal{D} > 2.8$, the segmentation begins, and stops when $\mathcal{D} < 1.2$, indicative of a pose ending.

2. *Filtering*: This step is used in order to eliminate some parts of the data stream that do not contribute to the gesture. The filtering process is used to eliminate portions of the data stream that do not contribute to the gesture. This work uses two low pass filters to reduce noise effects in data sets, and used the Bayesian decision rule to set the thresholds. If $\mathcal{D} < 1.4$, the data is not included in the pose data stream. Again, cross validation was used to verify these empirical result based on the data points obtained from a single subject.

3. *Classifier*: This step is used in order to identify the input gestures according to the database. Acceleration and gyroscopic measurements from the sensor boards are read in every time a pose is executed and compared to the pose signatures obtained from the training phase. The pose classifier during this supervised learning process is a minimum

mean distance rule classifier. Every pose is characterized by the mean and standard deviations of the components of its training feature vectors. The distance between an unknown sample M (input) and the mean of the features of class m (trained pose signature out of a total of L), $d(M, m)$, is then computed. The unknown input is then assigned to a class m^* (or move m^*) for which the distance, $d(M, m)$ is the minimum. Mathematically,

$$m^* = \min\{d(M, m)\},\ i = 1, 2, \ldots, L. \tag{2}$$

The distance, $d(M, m)$ is the Mahalanobis distance metric, and is based on correlations between variables by which different patterns can be identified and analyzed. When the correlation between the patterns decreases, then the Mahalanobis distance metric reduces to the more widely used Euclidean distance metric. In other words, the two distance metrics differ from one another in such a way that the Mahalanobis distance metric considers the correlations of the data set and is scale-invariant. The Mahalanobis distance metric between two vectors, $\mathbf{u}$ and $\mathbf{v}$ is given as follows:

$$d(\mathbf{u}, \mathbf{v}) = \sqrt{\sum \frac{(u_i - v_i)^2}{\sigma_i^2}}, \tag{3}$$

where σ_i is the standard deviation of the data over the sample set. If the distance falls within an acceptable threshold, the pose is declared as being detected. A high-level flowchart of this classifier is summarized in Fig. 4

4. Experimental Results

A series of experiments were performed to examine the capabilities and limitations of the two-sensor-based pose detection architectures in the mobile gaming context. The subjects executed a sequence of typical DDR steps multiple times consisting of 24 distinct poses for Experiment 1, and five distinct poses for Experiments 2 and 3. Tests were conducted under supervision with a consistent level of instruction and coaching given by a DDR professional.

4.1. Placement of Sensors

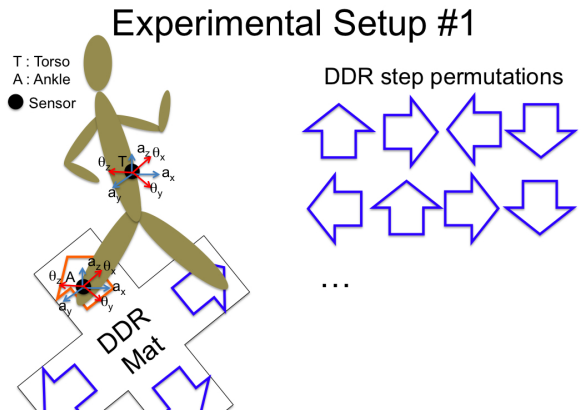
The accelerometer and gyro sensor boards were placed on the ankle and the torso during Experiments 1 (shown in Fig. 5(a)) and 3 (shown in Fig. 5(b). During Experiment 2, one of the sensor boards was placed on the palm instead of ankle (shown in Fig. 5(b)). Figs. 6(a)–6(c) show the standard deviations in the measurements in the accelerometer and gyro (all six dimensions) at the different locations during each experiment (multiple trials). The larger the standard deviation in the measurements, the higher is the probability for successful pose recognition. In other words, very miniscule standard deviation in measurements is indicative of a poor choice in the location of the sensor board on the human body. From Fig. 6(b), because the standard deviation in the accelerometer and gyro measurements is very low, it implies that placing the sensor board on the palm is ineffective for pose recognition. In other words, the set up in Experiment 2 reduces to just a single sensor board (on the ankle) as the one on the palm is redundant.

4.2. Learning Curve

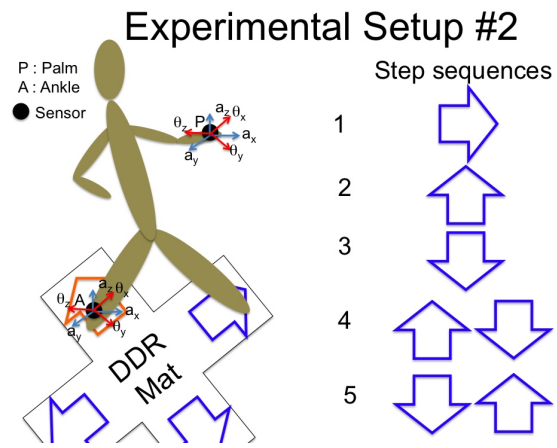
The learning curve experiment was conducted to examine the effects of repetition on the successful recognition rate of poses. The hypothesis was that with practice, any user would be able to achieve full recognition for every pose in a trial. In a mobile gaming context, we also want the players to progressively succeed with practice. Each participant performed the poses depending on the experiment. The resulting trend lines (for 5 out of the 48 distinct poses (shown in Fig. 7)) illustrated in Fig. 5(a), indicate that given enough repetition, any user can improve. It is interesting to note that over time, the participants' improvement would plateau as fatigue set in. Further experimentation is required to determine the optimal number of trials to maximize success. Although there was observed improvement overall, certain poses were problematic for most subjects, which led to further pose-based analysis discussed next.

4.3. Pose Difficulty

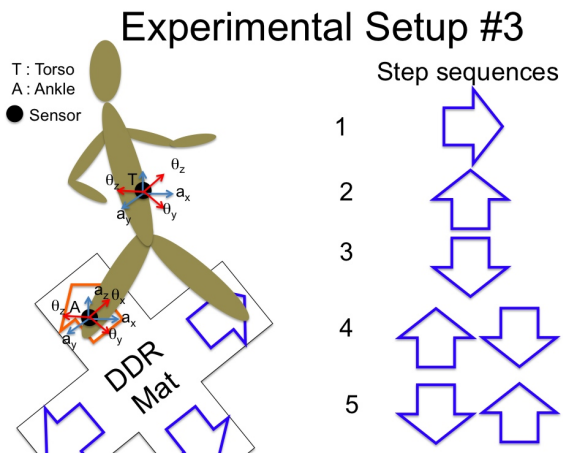
While performing the learning curve experiment, it was noted that certain poses were consistently problematic. Training data from a single individual was tested over 20 trials



(a) Sensor placements: One sensor board was strapped tightly to the ankle, while the other was strapped around the torso (or the waist).

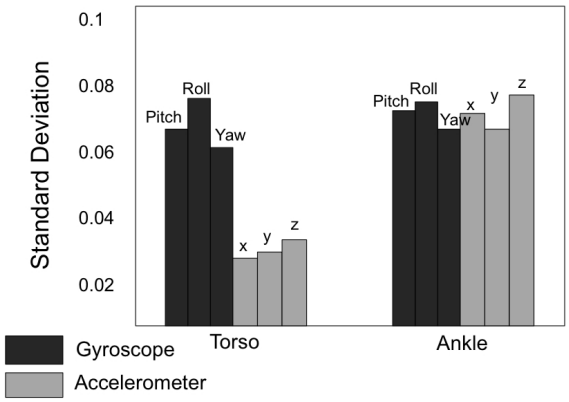


(b) Sensor placements: One sensor board was strapped tightly to the ankle, while the other was positioned on the palm.

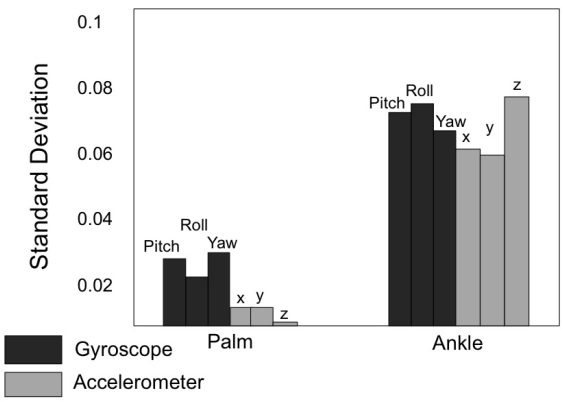


(c) Sensor placements: One sensor board was strapped tightly to the ankle, while the other was strapped around the torso (or the waist), just like the setup for Experiment 1, but the poses were different.

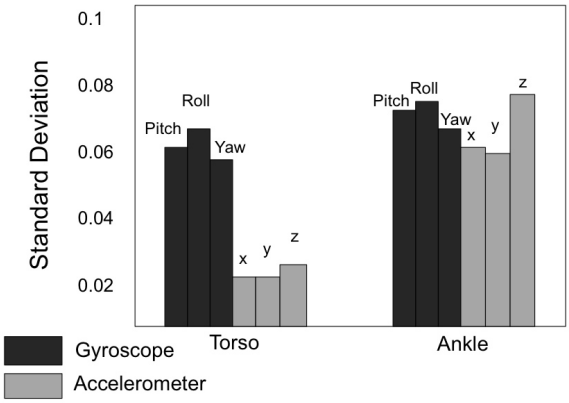
Figure 5: Sensor placements for the three DDR experiments.



(a) Experiment 1.



(b) Experiment 2.



(c) Experiment 3.

Figure 6: Standard deviations in the measurements of the motion sensors over multiple trials for a single user.

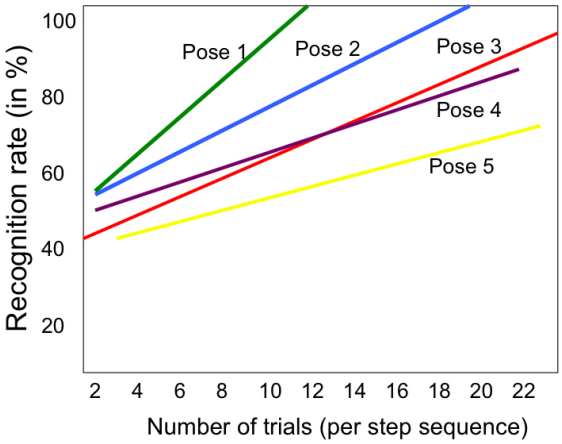
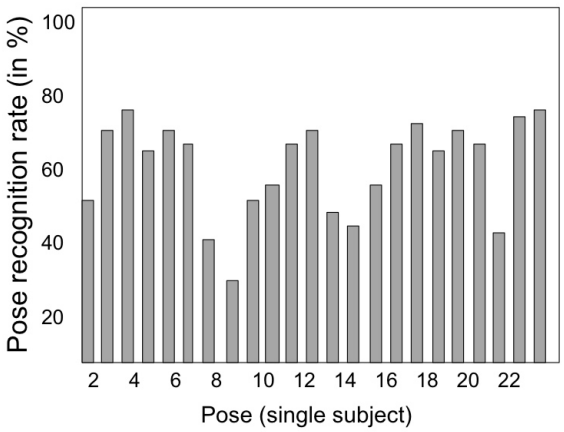
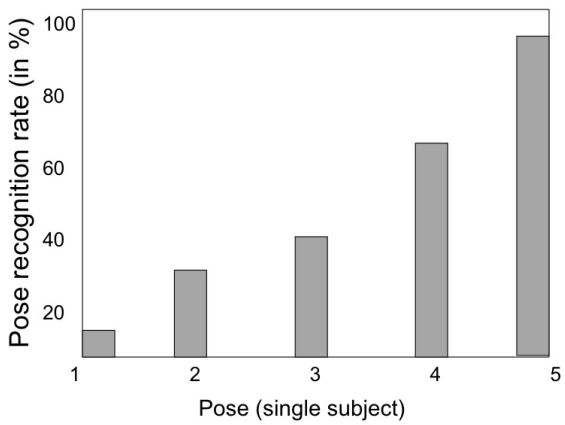


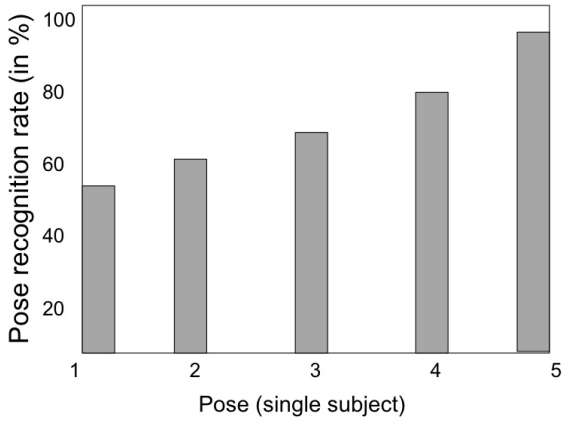
Figure 7: Trend lines to examine the effects of repetition for a single subject.



(a) Experiment 1.



(b) Experiment 2.



(c) Experiment 3.

Figure 8: Pose recognition rate for a single subject.

using a low recognition threshold determined by continually increasing the allowable recognition threshold in order to determine the level of difficulty among the different poses for the experiments. The graph in Figs. 8(a), 8(b), and 8(c) illustrate a visible difference in the recognition rates for the poses during Experiments 1, 2, and 3, respectively. For example, in Fig. 8(a), Pose 8 required extreme movements (i.e., forward-backward-left-right), which can be difficult and had a recognition rate of about 28%, a little less compared to other poses. However, with sufficient training and practice, the recognition rates can be enhanced.

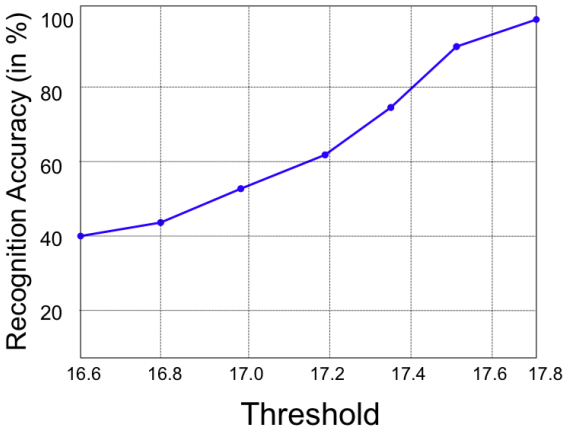
Although it is desirable for the population to be capable of performing every pose, a difference in complexity among the poses can prove to be beneficial in some aspects of

game development, allowing for an added level of difficulty in the game. The results from the experiments suggest a need for a larger training data set for the poses in order to accommodate a wider variety of body types and physical limitations in addition to expertise levels. This should not be surprising, because it shows that a large training sample is not necessarily required to create games that are playable by a range of people.

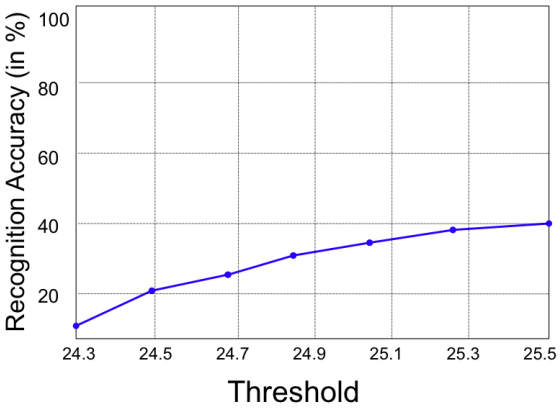
4.4. Threshold

It is unreasonable to expect a user to be able to achieve a "perfect" pose, in other words achieve a recognition distance of zero, even if they are a part of the training sample. Simple variations are inevitable due to the core mechanics of the human musculature system. Therefore, an allowable range, defined by a threshold, is experimentally determined to allow reasonably similar poses to be detected while avoiding false alarms. The threshold for an individuals data set was determined by continually increasing the allowable Mahalanobis distance after each trial until the lowest possible value is reached where all user poses could be recognized. It should be noted that the threshold for one individual data set will not necessarily be the same as anothers since their standard deviations will vary based on how they perform the pose in the training phase and how steady they are during recording.

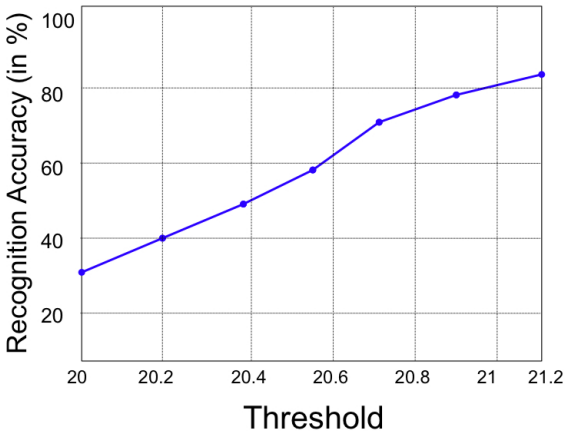
From Figs. 9(a) and 9(c), it can be observed that thresholds of 17.8 and 21.2 are acceptable because they achieve a pose recognition rate of close to 98% and 80% for Experiments 1 and 3, respectively. Figs. 9(a) shows the success rates and thresholds for a single user for Pose 8. It is believed that with more training these recognition rates will go up without having to increase the thresholds too much. For Experiment 2, the recognition rate has a monotonic upward trend, however, it plateaus out eventually at about 40%. This implies that using just one sensor (because the sensor on the palm does not experience lateral or angular displacements) is not sufficient to capture the dynamics that are needed for pose recognition. Once an acceptable threshold was found of 17.8 for Experiment 1, a confusion matrix was created to determine the likelihood of false positives among other poses. The results indicate little possibility for the occurrence of false positives.



(a) Experiment 1.



(b) Experiment 2.



(c) Experiment 3.

Figure 9: Success rates and thresholds for a single user.

4.5. Training and Standard Deviation

The standard deviations increase with the extent of variation allowed in a pose during the training. Understandably, this makes the pose more easily detected. If the deviation gets too high, more error and thus less precise poses will be tolerated. Contrarily, if the standard deviation is too low, it will be very difficult to detect the pose. Standard deviations vary based on several factors: the person recording the pose (e.g., size, shape, athletic-built, energy levels of the subject, attire), the complexity of the pose, amount of practice (expertise level) and the recording time. If the person recording the pose is unable to consistently maintain a pose, standard deviation may be unusually high, whereas if they remain perfectly still, it will be low. For this reason, in our experiments, the subjects perform each pose 20 times after some initial practice, with breaks in between to ensure some variation. Similarly, poses that require balancing often result in higher deviations.

It became apparent throughout our experiments that even with practice, the same subject was unable to successfully replicate certain poses from the training data set. Going forward, it is likely that some other subjects might not be able to replicate certain poses from a single individual's training data set. Physical limitations can make it almost impossible for some subjects to get their body into the exact same position as the person who trained the date set. These restrictions include things like muscle size, limb length, proportions, body type, athletic ability, expertise, and their enjoying dancing in general. Therefore, combining multiple peoples training data sets would make it possible for a variety of users to achieve higher recognition rates.

5. Conclusions

In this work, it has been shown that using a combination of six sensors (three accelerometers and three gyros), one can detect and discern human dance movements pertaining to a DDR-like game with a reasonably high probability. These initial results obtained using two sensor boards with the form factor of mobile phones, but without the processor serves as a motivation for future research in mobile gaming using dual mode design phones. We

have created a relatively inexpensive way of accurately identifying specific poses for gaming on a mobile platform. Though a new user may not achieve 100% pose recognition, it has been shown that with sufficient practice, improvement will be seen. The effect is desirable in game development to help maintain user interest over time. Furthermore, recognition rates continue to improve as a larger and more diverse sample of training data is used. Game complexity can be naturally introduced by playing with the threshold value for recognition allowing easy, normal, hard, and very tough levels of difficulty/expertise.

References

[1] E. M. I. Koivisto, "Mobile games 2010," *Proc. International Conference on Game Research and Development*, 2006, pp. 1–2.

[2] Avalaible: `http://www.gartner.com/it/page.jsp?id=507467`

[3] J. O. B. Soh and B. C. Y. Tan, "Mobile gaming," *Commun. ACM*, vol. 51, no. 3, pp. 35–39, 2008.

[4] A. V. Nasrulloh, T. L. R. Mengko, and T. Apriantono, "Development of dual video acquisition and parallel AVI player with camera calibration system for supporting dual view motion detection," *Intl. J. of E-Health and Medical Communications (IJEHMC)*, vol. 1, no. 3, pp. 41–49, 2010.

[5] N. Noury, *et. al.*, "A MEMS-based microsystem for the monitoring of the activity of frail elderly in their daily life: The actidom project," *Proc. Sixth International Workshop on Enterprise Networking and Computing in Healthcare Industry*, 2004, pp. 105–109.

[6] A. Wixted, D. Theil, A. Hahn, C. Gore, D. Pyne, and D. James, "Measurement of energy expenditure in elite athletes using MEMS-based triaxial accelerometers," *IEEE Sensors J.*, pp. 481–488, 2007.

[7] R. Zhu and Z. Zhou, "A real-time articulated human motion tracking using tri-axis

inertial/magnetic sensors package," *IEEE Trans. Neural Syst. Rehabil. Eng.*, pp. 295–302, 2004.

[8] P. H. Veltink and H. M. Franken, "Detection of knee unlock during stance by accelerometry," *IEEE Trans. Rehabil. Eng.*, pp. 395–402, 1996.

[9] I. J. Jang and W. B. Park, "Signal processing of the accelerometer for gesture awareness on handheld devices," *Proc. 12th IEEE International Workshop on Robot and Human Interactive Communication*, Oct.-Nov. 2003, pp. 139–143.

[10] H. J. Luinge and P. H. Veltink, "Inclination measurement of human movement using a 3-D accelerometer with autocalibration," *IEEE Trans. Neural Syst. Rehabil. Eng.*, pp. 112–121, 2004.

[11] L. Hoff, O. Elle, M. Grimmes, S. Halvorsen, H. Alker, and E. Fosse, "Measurements of heart motion using accelerometers," *Proc. 26th Annual International Conference of the IEEE EMBS*, 2004, pp. 2049–2051.

[12] R. Zhang, M. Loschonsky, and L. M. Reindl, "Study of zero velocity update for both low- and high-speed human activities," *Intl. J. of E-Health and Medical Communications (IJEHMC)*, vol. 2, no. 2, pp. 46–67, 2011.

[13] H. J. Luinge, P. Veltink, and C. Baten, "Estimation of orientation with gyroscopes and accelerometers," *First Joint BMES/EMBS Conference Serving Humanity, Advancing Technology*, 1999, pp. 844.

[14] K. Motoi, S. Tanaka, Y. W. Jang, and H. C. Kim, "Evaluation of a new sensor system for ambulatory monitoring of human posture and walking speed using accelerometers and gyroscope," *Proc. SICE Annual Conference in Fukui*, 2003, pp. 1223–1235.

[15] M. F. Nowlan, "Human recognition via gait identification using accelerometer gyro forces." Available: `cs-www.cs.yale.edu/homes/mfn3/proj/mfn_gait_id.pdf`

[16] N. Bradai, L. Chaari, and L. Kamoun, "A comprehensive overview of wireless body area networks (WBAN)," *Intl. J. of E-Health and Medical Communications (IJEHMC)*, vol. 2, no. 3, pp. 1–30, 2011.

[17] A. Fleury, N. Noury, and M. Vacher, "Improving supervised classification of activities of daily living using prior knowledge," *Intl. J. of E-Health and Medical Communications (IJEHMC)*, vol. 2, no. 1, pp. 17–34, 2011.

[18] T. Iso and K. Yamazaki, "Gait analyzer based on cell phone with a single three-axis accelerometer," *Proc. 8th Conference on Human-Computer Interaction with Mobile Devices and Services*, 2006, pp. 141–144.

[19] T. Iso and K. Yamazaki, "Extracting personal user context with a three-axis sensor mounted on a freely carried cell phone," *Proc. International Conference on Signal Processing and Multimedia Applications*, 2006, pp. 224–231.

[20] D. Karantonis, M. Narayanan, M. Mathie, N. Lovell, and B. Cellar, "Implementation of a real-time human movement classifier sing a triaxial accelerometer for ambulatory monitoring," *IEEE Trans. Inf. Technol. Biomed.*, pp. 156–167, 2006.

[21] K. -T. Song and C. -W. Chen, "Multi-person pose recognition using a zigbee sensor network," *Proc. 11th World Congress the International Federation of Automatic Control*, Jul. 2008, pp. 14976–14981.

[22] N. Crampton, K. Fox, H. Johnston, and A. Whitehead, "Dance, dance, evolution: Accelerometer sensor networks as input to video games," *Proc. IEEE International Workshop on Haptic Audio Visual Environments and their Applications*, Oct. 2007, pp. 107–112.

[23] Source: `http://www.invensense.com/`