Toward a Non-Intrusive, Physio-Behavioral Biometric for Smartphones

Esther Vasiete
Department of Computer Science
University of Colorado Boulder
esther.vasiete@colorado.edu

Yan Chen
Department of Computer Science
University of Colorado Boulder
yan.chen@colorado.edu

Ian Char
Department of Computer Science
University of Colorado Boulder
ian.char@colorado.edu

Tom Yeh
Department of Computer Science
University of Colorado Boulder
tom.yeh@colorado.edu

Vishal Patel
UMIACS
University of Maryland
College Park
pvishalm@umiacs.umd.edu

Larry Davis
UMIACS
University of Maryland
College Park
lsd@umiacs.umd.edu

Rama Chellappa
UMIACS
University of Maryland
College Park
rama@umiacs.umd.edu

MobileHCI '14, Sep 23–26 2014, Toronto, ON, Canada
ACM 978-1-4503-3004-6/14/09.
http://dx.doi.org/10.1145/2628363.2634223

Abstract
Biometric authentication relies on an individual's inner characteristics and traits. We propose an active authentication system on a mobile device that relies on two biometric modalities: 3D gestures and face recognition. The novelty of our approach is to combine 3D gesture and face recognition in a non-intrusive and unconstrained environment; the active authentication system is running in the background while the user is performing his/her main task.

Author Keywords
3D gesture recognition; behavioral biometric authentication; face recognition; fusion system.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
The most common user identification techniques on mobile devices are based on passwords or on graphical puzzles. However, there are three shortcomings with these techniques. First, passwords in memory-based approaches can be forgotten, stolen, or hacked. Second, after the initial login there is no further identification procedure as long as the device remains active. An impostor could gain physical access to the device if the legitimate user leaves it unlocked and unattended. Third, the authentication mechanism is

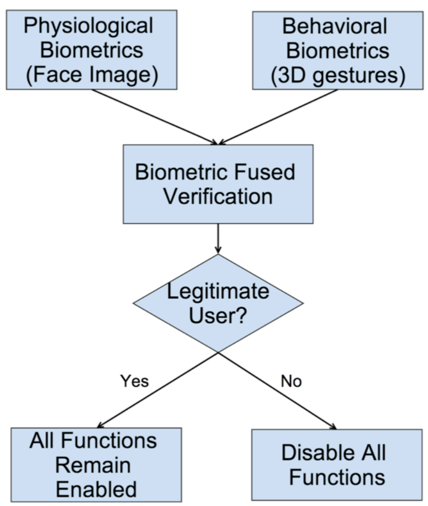

Figure 1. Proposed physio-behavioral biometric for non-intrusive active authentication.

often intrusive and disruptive; users can be momentarily forced away from their main task to enter passwords and/or perform gestures.

We propose a new approach that can address these three shortcomings (Fig. 1). Our approach is behavior-based (rather than memory-based), active (rather than login-only), and, most importantly, non-intrusive. The property of being non-intrusive is valuable from a mobile HCI perspective, as it would greatly increase the usability of the authentication procedure. On the technical side, our approach will fuse physiological and behavioral biometrics. Biometric authentication relies on an individual's personal characteristics and traits that cannot be forgotten and could be difficult to steal or mimic.

Today's smartphones come with a growing number of embedded sensors such as an accelerometer, gyroscope, digital compass, front and rear cameras, and ambient light and proximity sensors. We take advantage of the variety of available sensors to fuse two forms of biometrics: face recognition and behavioral movement dynamics. This fusion creates a unique signature that is constantly being extracted and analyzed to determine whether the current user matches the previously stored signature for the authorized user. If the current biometric measurements do not match the reference signature, the mobile device can lock itself to prevent further unauthorized use.

Other highly reliable forms of biometric authentication exist, such as fingerprint identification and retinal or iris scans, but these require the subject's participation and the use of external, often expensive, equipment. Our proposed biometric fusion authentication technique is non-intrusive (face recognition and behavioral movement analysis is performed in the background while users are performing their main tasks) and does not require any extra equipment—only the existing sensors are needed.

Related Work

Accelerometer and gyroscope sensors have been exploited extensively to capture mobile users' behavioral characteristics. In [1], a set of 43 generated features from 10-second interval accelerometer readings produced a positive authentication rate of 82.1%-92.9%. These features were independent of the user state (walking, jogging, going upstairs or downstairs). Wolff [2] also studied the accelerometer dynamics by extracting the variance in acceleration and orientation across the three dimensions (x, y, and z) and used a Gaussian distribution model to achieve a user classification accuracy of 83%. In [3], researchers used accelerometer and gyroscope readings during screen taps and an ensemble of machine learning classifiers to predict the letters and icons that were pressed with an accuracy of up to 80% and 90%, respectively. These findings further justify the need to increase security systems on our mobile devices.

The main drawback of 3D gesture recognition is that it is highly task-dependent and authentication is achieved by means of knowing the task context. Face recognition, on the other hand, does not require context information although it requires a partial face to be captured by the camera.

The utilization of almost real-time face recognition techniques in mobile devices is now feasible due to the increases in processing power [4, 5, 6]. To date, biometric authentication in mobile devices is used for

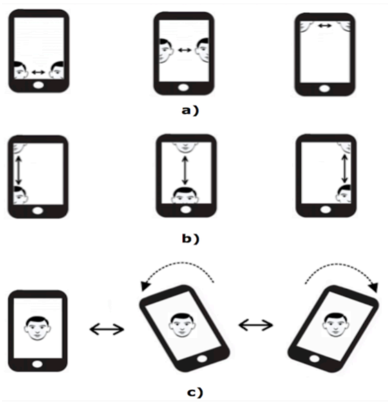

Figure 2. Enrollment task. Users are asked to scan their faces (a) horizontally, (b) vertically, and, (c) rotationally.

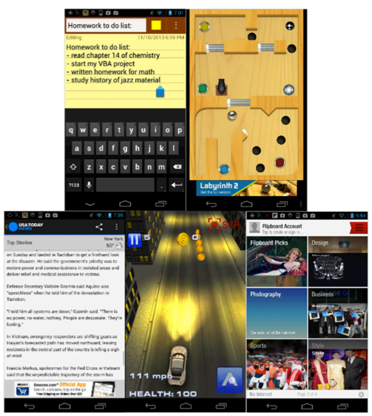

Figure 3. Snapshot from each of the five tasks performed by 30 subjects while the front-facing camera is *non-intrusively* capturing in the background.

access control (rather than a PIN or password) that corresponds to a highly constrained scenario. The combination of more than one biometric modality can increase the robustness of the authentication system. A combination of face and speech recognition in [7] outperforms either modality on its own and provides a performance gain of more than 25%. However, these existing authentication methods are still conducted intrusively during login time.

The novelty of our approach is to combine 3D gesture and face recognition in a non-intrusive and non-constrained environment—the active authentication system runs in the background while the user performs any other task (Fig. 1). Furthermore, sensor data such as proximity and ambient light will serve as inputs to the face recognition system to provide a novel way of dealing with pose and illumination variation, which still constitute the main issues in unconstrained face recognition [8, 9, 10].

Increasing Robustness

Previous work shows promises of physio-behavioral biometric data for authentication under constraint scenarios, but there are still diverse technical challenges and problems when dealing with non-intrusive authentication. Here, we outline our approach to providing robustness by combining state-of-the-art face recognition and 3D gesture recognition methods.

Face Recognition

Holistic approaches that use the whole face region as the raw data have received substantial attention among all face recognition studies in recent years [11]. Turk and Pentland [12] used principal component analysis (PCA) for face recognition by projecting face images onto a dimensionality-reduced feature space, creating the so-called Eigenfaces. Other feature spaces such as linear discriminant analysis (LDA) and independent component analysis (ICA) have also been successful. However, recent work has shown that PCA performs better than LDA and ICA when the correct distance metric is used [13, 14].

Recognition performance drastically drops when illumination and pose variations in face images are encountered. Significant work has been done to combat these issues [9, 10]. Other researchers have incorporated the extraction of key visual features to gain robustness. Instead, we propose the incorporation of sensorial features to compensate for changes in the illumination and to perform pose inference.

The accelerometer and gyroscope sensors were used in [15] to infer and compensate tilted images for face detection. We propose using the sensors that follow:

- Proximity sensors measure the proximity of an object relative to the device screen They can help infer when a user is placing a call (images will be discarded) or when the user is tapping the screen (face might be occluded by a finger).
- Light sensors measure the illumination level. Knowing the real lighting conditions makes illumination compensation technique more accurate and efficient.
- Orientation sensors measure the device’s rotation around all three physical axes (x, y, z). These are used for tilt compensation in order to provide an image with an upright face to facilitate the face detection task and pose inference for view-based face recognition. Appearances can change depending on the subject’s position with respect to the phone’s orientation.
- Motion sensors such as the accelerometer, gyroscope, and gravity sensors will ease tracking of the

face. For instance, intensity of motion can be used to discard images too blurry due to excessive vibration.

3D Gesture Recognition

Subjects may be authenticated through gestures when using their mobiles. Motion sensors can be used to analyze each person's movements or behavior. Although hierarchical probabilistic models have been popular for human behavior modeling, they could lead to over-fitting when dealing with sensor data over long periods [16], especially in real scenarios where abnormal behavior is usually present. Recent work has used supervised or unsupervised machine learning techniques for authentication under the smartphone platform. A Gaussian mixture model is used to model pedestrians' behavioral trajectories in [17]. A user can be authenticated in [1] while walking, jogging, or climbing stairs with a phone inside a pocket. In active authentication, it is preferable to provide authentication security while the mobile device is being used. In [18], touch screen data such as pressure, size, and speed are analyzed to add an extra security layer to password patterns. Pinch and spread gestures have been found to fit a Fitts's law model on a specific resizing task [19].

Until now, behavioral authentication systems on mobile devices have been task-specific. It is important to acknowledge that 3D gestures vary greatly depending on the task (e.g., web browsing and playing a car racing game show two distinct movement dynamics). Rather than task-specific, we build models that are gesture-specific by grouping similar actions into four main classes:

1. Low-motion gestures such as carrying the phone while reading the news, checking social networks, or watching a video.
2. High-motion gestures often present when playing accelerometer games.
3. Finger-tapping. Samples can be extracted from typing, icon selection, and similar actions.
4. Swiping gestures, during scrolling, switching windows, or playing some games.

The main challenge is to extract features for each of the four classes that will retain the most within-subject variability while discriminating between-subject variability; then, a classifier will be built to provide good predictions. By aggregating time series data into samples, we compute Fourier components, moments, mean acceleration per axis, and other information.

Fused System

We acknowledge that the noisy measurements, high variance of data, and difficulties derived from an unconstrained scenario (in which the user interacts with her device with total freedom) will cause authentication accuracy to drop in any modality; however, we can expect sufficiently high authentication accuracy when combining both modalities. To accomplish this, we represent new data as a sparse linear combination of concatenated features from two or more modalities [20]. Late fusion can also be applied by combining prediction scores from both modalities into a simple score: the fused score. Any of these solutions is expected to provide higher authentication accuracy than any single modality on its own.

Preliminary Study

Our main research question is to understand to what extent behavioral biometric data collected *without* user cooperation in a non-intrusive manner can capture a user's unique characteristics for authentication. For instance, can the front-facing camera capture enough of a user's face for reliable face recognition when the

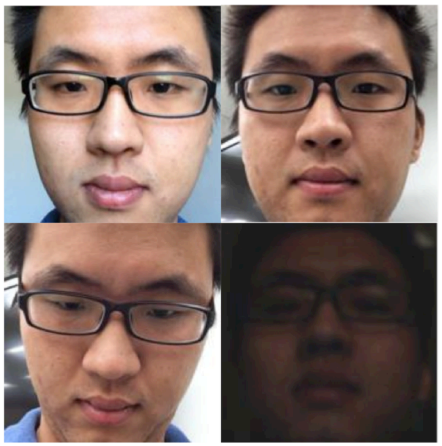

Figure 4: User appearance changes due to different pose and illumination conditions.

Figure 5. Some images extracted from the front camera. Note variation in illumination, pose, expression and face size.

user is performing a main task (e.g., playing a game) rather than when being prompted?

Data Collection

To this end, we collected a dataset of 30 subjects (20 males and 10 females). Our data collection was designed to monitor as many biometric qualities of a subject as possible while the subject performed a number of tasks (Fig. 3). We controlled for the device variation by providing each subject a Nexus 4 phone. The subjects were told to spend 5 minutes performing each of the following tasks: use a text editor application and write a response to a question that we chose, read news articles on the *USA Today* application, browse articles and pictures using *Flipboard*, play a ball balancing game called *Labyrinth*, and play a racecar game called *Crime Racing City*. Subjects performed these tasks with no further constraints (e.g., they could read whatever articles interested them and hold the phone however they wanted to). While the subjects were completing these tasks, three applications were collecting data in the background: one used the front camera to capture video of the subject's face; one recorded what was being displayed on the screen; and the last one recorded a series of sensor data. This series of sensor features included an accelerometer, a gyroscope, the device's orientation, linear acceleration, magnetic field and gravity across three dimensions, and ambient light, proximity, location, temperature, and air pressure.

In addition, we included an enrollment task that required the subjects to hold the phone at different angles from their faces (Fig. 2). This enrollment task is cooperative and intrusive, and the data can be used as a baseline as well as training data to extract face images from different views and synchronize them with the device's position using the sensor data.

Challenges

Because of the nature of the front camera recording, there are several inherent problems that arise with facial recognition. One of the problems comes from how the camera is positioned, which can cause a huge difference in the subject's appearance (see Fig. 4). Other problems include partial faces due to the closeness of the phone, partial occlusion by a finger tapping the screen, or poor lighting, especially when the user leans over the phone and blocks out light sources. Fig. 5 demonstrates some of the aforementioned scenarios.

Baseline Performance Results

So far we have experimented with Eigenfaces for face recognition and Gaussian mixture models for 3D gesture recognition. We obtained 98% face recognition accuracy when selecting face images with no or little variation in pose and illumination. This number drops significantly when no human selection is made on the face images. For instance, another experiment consisted of performing Eigenfaces (keeping 96% of the variance) on face images detected by a cascade classifier during the enrollment task and performing classification on face images detected from the remaining tasks. We used a k-nearest neighbor for classification with a k value from 1 to 100 to build the receiver-operating characteristic (ROC) in Fig. 6. These results are promising taking into account the challenges of our dataset.

We also experimented with the sensor data to explore the potential of 3D gesture recognition. By extracting only three features—acceleration on the x axis,

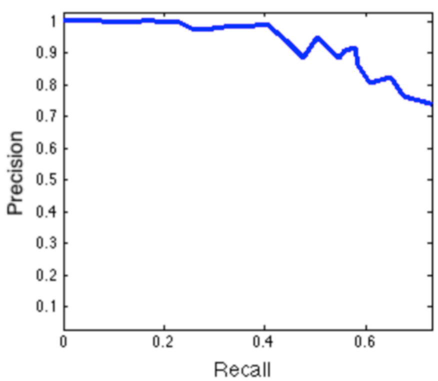

Figure 6. ROC for face recognition with Eigenfaces (face images during enrollment were used for training and face images from the remaining five tasks were used for testing).

gyroscope on the *y* axis, and orientation on the *y*—and modeling them using a Gaussian mixture model, we achieved 87% prediction accuracy when comparing the three features over the test data set for two different tasks. We aim to develop more robust models and take advantage of all the sensor data to provide features that will serve to authenticate the user under any task.

This work was supported by a DARPA grant FA8750-13-2-0279.

References

[1] Kwapisz, J.R., Weiss, G.M., and Moore, S.A. Cell phone-based biometric identification. *Biometrics*, 2010.

[2] Wolff, Matt. Behavioral Biometric Identification on Mobile Devices. *Foundations of Augmented Cognition*. Springer Berlin Heidelberg, 783-791, 2013.

[3] E. Miluzzo, A. Varshavsky, S. Balakrishnan, R. Choudhury. TapPrints: Your Finger Taps Have Fingerprints. *Proc. Mobile Systems, applications and services,* 2012.

[4] Yi-Chu Wang, Kwang-Ting Cheng. Energy-Optimized Mapping of Application to Smartphone Platform—A Case Study of Mobile Face Recognition. *ECVW*, 2011.

[5] K. Choi, K.-A. Toh, H. Byun. Realtime training on mobile devices for face recognition applications. *Pattern Recognition 44(2)*: 386-400, 2011.

[6] In-ho Choi, Kyoung-Sic Cho. Robust Facial Expression Recognition Using a Smartphone Working against Illumination Variation. *Applied Mathematics and Information Science*, No. 6, 403-408, 2012.

[7] C. McCool, S. Marcel, A. Hadid, M. Pietikainen, P. Matejka, J. Cernocky, N. Poh, J. Kittler, A. Larcher, C. Levy, D. Matrouf, J. Bonastre, P. Tresadern and T. Cootes. Bi-Modal Person Recognition on a Mobile Phone: using mobile phone data. *Proc. Multimedia and Expo Workshops*, 635-640, 2012.

[8] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld. Face Recognition: A Literature Survey. *ACM Computing Surveys*, 399-458, 2003.

[9] Xuan Zou; Kittler, J.; Messer, K. Illumination Invariant Face Recognition: A Survey. *Biometrics: Theory, Applications, and Systems*, 2007.

[10] Zhang, X., Gao, Y. Face recognition across pose: A review. *Pattern Recognition* 42(11): 2876-2896, 2009.

[11] R. Jafri, H. R. Arabnia. A Survey of Face Recognition Techniques. *Journal of Information Processing Systems*, Vol. 5, No. 2, 2009.

[12] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 71-86, 1991.

[13] Hyunjong Cho, Seungbin Moon. Comparison of PCA and LDA based face recognition algorithms under illumination variations. *ICROS-SICE*, 2009.

[14] K. Baek, B.A. Draper, J.R. Beveridge, K. She. PCA vs. ICA: A comparison on the FERET data set. *Joint Conference on Information Sciences*, 2002.

[15] Yang, X., You, C. W., Lu, H., Lin, M., Lane, N. D., Campbell, A. T. Visage: A Face Interpretation Engine for Smartphone Applications. *Proc. MobiCASE*, 2012.

[16] Atallah, L., Yang, G.Z. The use of pervasive sensing for behaviour profiling—a survey. *Pervasive Mob. Comput.* 5(5), 447–464, 2009.

[17] Neil Johnson, David Hogg. Representation and synthesis of behaviour using Gaussian mixtures. 20(12), 889–894, 2002.

[18] Alexander De Luca, Alina Hang, Frederik Brudy, Christian Lindner, Heinrich Hussmann. Touch me once and I know it's you! Implicit authentication based on touch screen patterns. *CHI*, 987-997, 2012.

[19] J. Tran, S. Trewin, C. Swart, B.E. John, J.C. Thomas. Exploring pinch and spread gestures on mobile devices. *MobileHCI '13,* 2013.

[20] Sumit Shekhar, Vishal M. Patel, Nasser M. Nasrabadi, Rama Chellappa. Joint Sparse Representation for Robust Multimodal Biometrics Recognition. *IEEE PAMI.* 36(1), 113-126, 2014.