An Approach for User Identification for Head-Mounted Displays*

Cynthia E. Rogers, Alexander W. Witt, Alexander D. Solomon, Krishna K. Venkatasubramanian†
Department of Computer Science, Worcester Polytechnic Institute
100 Institute Road, Worcester, MA, USA, 01609
{cerogers, awitt, adsolomon, kven}@wpi.edu

ABSTRACT
A head-mounted display (HMD) is a device, worn by a person, which has a display in front of one or both eyes. HMDs have applications in a variety of domains including gaming, virtual reality, and medicine. In this paper we present an approach that can identify a user, from among a group of users, by synchronously capturing their unconscious blinking and head-movements using integrated HMD sensors. We ask *each user* of the HMD to view a series of rapidly changing images of numbers and letters on the HMD display. Simultaneously, their blinks and head-movements are captured using infrared, accelerometer, and gyroscope sensors. Analysis of our approach using blink and head-movement data collected from 20 individuals demonstrates the feasibility of our approach with an accuracy of $\sim$94%.

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces - Input devices and strategies

Author Keywords
Wearable Devices; Head-Mounted Displays; User Identification; Blinks; Head-Movement

INTRODUCTION
A head-mounted display (HMD) is a device, worn by a person, which has a display in front of one or both eyes. HMDs come in many forms, including those that project images and those that have transparent viewports. In recent years, HMDs have been developed for a variety of applications such as virtual reality (Agevant Glyph - http://www.avegant.com), gaming (Oculus Rift - https://www.oculus.com/rift/), general-purpose displays (Google Glass - http://www.google.com/glass/start).Thus far the focus has been on demonstrating technical feasibility of the technology, however there are still several important issues that need to be addressed. These include access control (often for privacy protection), ease of interaction, and personalization [4]. Support for each of these requires the capability of - *user identification*.

User identification enables HMDs to recognize users who are interacting with (i.e., wearing) them. Formally, given a set of N users of a particular HMD, the task of user identification is to determine which of the N users is currently using (i.e., wearing) the device. The N users of the HMD in this context are known *a priori*, and at any given time the HMD is worn by one of the N users.

Traditionally, user identification in the wearable computing domain, which encompasses HMDs, has been performed in one of three general ways. (1) Requiring devices to be statically associated with a given user [3]. This association is accomplished by pairing a wearable device (e.g., Google Glass) with the user's smartphone or console (e.g., XBox). (2) Asking the user to input their id, pin, pattern, or password for identification. (3) Using biometrics such as voice or face recognition to identify the user. In this paper we present a new and complementary solution that can enable user identification to occur in a manner that is minimally intrusive because it does not require the user to engage in any deliberate activity. The minimally intrusive nature of our solution is important because it accords to the demand for increased usability on HMDs [4].

At this stage we would like to clarify that *user identification is not the same as user authentication*. The goal here is to identify a user from among a group of known users. Authentication focuses on independently identifying a user from any other user who is likely to use the system. In this preliminary work, we focus on the relatively simpler problem of user identification. This means the designed approach should meet two *design goals*: (1) it should be able to automatically identify who is wearing the HMD from among a group of users; and (2) it should not require users to input any identifying information because doing so can limit HMD usability. Given our design goals, we aim to design an approach that can identify a user by synchronously capturing their unconscious blinking and head-movements.

To demonstrate the feasibility of our approach for user identification in HMDs, we collected data from 20 different individuals using Google Glass as an HMD platform. The preliminary analysis of our approach shows a user identification accuracy of over 94%. Overall, we demonstrate the *feasibility* of using blinking and head movement for identifying users wearing an HMD, in a minimally intrusive manner, without requiring

*The first three authors are undergraduate students who performed this work as part of their final year project.
†Point of Contact

ISWC '15, September 7–11, 2015, Osaka, Japan.
Copyright 2015 © ACM ISBN 978-1-4503-3578-2/15/09 . . . $15.00.
http://dx.doi.org/10.1145/2802083.2808391

explicit user input. In the rest of the paper we use the terms user and subjects interchangeably.

RELATED WORK

The use of blinking as a biometric for user identification is not necessarily new. Westeyn and Starner [9] propose the use of deliberate blinking to song cadences as a way to identify users. The song cadences act as passwords, which if correctly "played" enables user identification. The approach utilized inter-blink interval, and duration of eye closure as the primary features of interest and achieved a 82.02% accuracy in identifying a user from among nine users when all users were asked to blink the same pattern. The authors therefore concluded that blinking could be used as an additional biometric in user authentication. Further, the approach relies on deliberate blinking, something that we intend to avoid in our approach. In [1, 2], the authors propose a mechanism for obtaining an individual's blinking patterns using brain waves, and they demonstrate high identification accuracy. However, both of these approaches require the ability to measure a person's EEG, a capability that most HMDs do not possess. Head movement has also been suggested as a way to identify individuals [6], however, not much work has been done in this regard. In [7], the authors present an approach that uses an HMD to detect blinking behavior and head movement to determine what task the user is performing. While this research is useful, it deviates from our objective of user identification. Its primary concern is to identify types of actions performed by a user. In summary, the work done thus far has not been sufficient to demonstrate a high accuracy, minimally intrusive user identification on HMDs.

APPROACH

We use Google Glass as our HMD platform. The principal reasons being: (1) it is an excellent model for the general technical capabilities of HMDs (even though it is designed for individual use and not for sharing); (2) it was relatively easy to procure; and (3) it has the infrared, accelerometer and gyroscope sensors needed to implement our approach. In this section we describe the four operational stages of our approach in more detail.

Data Collection

We use an infrared (IR) sensor built into the HMD platform (Google Glass) to detect the user's blinking pattern. We directly access the IR sensor built into the Google Glass platform to obtain and detect the user's blinking pattern. The IR sensor in the Google Glass performs multiple distance measurements between the sensor and the eye. The distance of a user's eye in relation to the IR sensor decreases when their eyelid is closed and increases when their eyelid is open. With respect to head-movement, we use the output of the accelerometer and gyroscope sensors in the HMD to detect head-movement in three dimensions. We use the gyroscope sensor to differentiate between types of head-movement observed from users, particularly the direction of movement. The accelerometer sensor provides us a way to determine the magnitude of a given head-movement. These accelerometer and gyroscope sensors provide us with two separate time series.

The data from the IR, accelerometer, and gyroscope sensors are collected from users while a rapid sequence of images (with a single letter or number on them) are shown on the HMD display. The use of images provides (1) visual fixation for the users during the data collection/identification process, and (2) standardizes the duration of data collection across users. The use of the image sequence precludes the need for the user to enter any identifying information. The images trigger induced blinks (IBs) in the viewer. IBs are triggered when a user is shown a visual stimulus, particularly during the explicit and implicit shifts in the content [8].

Figure 1 shows *an exact* illustration of the sequence of images as they are visible to the users. When users put on the HMD, they are shown an initial sequence of instruction slides, which inform them to focus on the numbers and letters that will soon appear. The instruction sequence is then followed by an audible cue and an image of a plus sign on the HMD display, which informs the user that the image sequence is about to start. Ten seconds later the image sequence commences. After the initial instructions have been displayed, we start monitoring the user's blinks. While the plus sign is being shown, the blinks are essentially spontaneous blinking (SB) (i.e., unconscious blinking to maintain the tear-film in the eye as opposed to IB which is more deliberate). We refer to this stage of blink data collection as the **SB phase**. Once the rapidly changing sequence of images starts playing, we refer to this as the **IB phase** of blink data collection, as we are most likely observing induced blinks.

The rapidly changing sequence of images is shown to each of the users multiple times during the data collection process in order to capture the variation in their blinking and head-movement. Each cycle of the image sequence that is shown to the user is called **image sequence iteration**. In our experiments the image sequence iteration was 34 seconds long. We chose such a long duration to be able to capture the individual variations in the user blinking and head movement effectively.

Feature Extraction

Once the sensor readings are collected over several image sequence iterations, the next step is to extract features from them that can then be used to identify the user. This section describes the different features extracted from the raw readings from the IR, accelerometer, and gyroscope sensors in the HMD. From each image sequence iteration we extracted a total of 162 features from the three sensor readings to create a feature point (vector) for a user. We do not list all of the features here due to limited space. Instead we present a categorization of the features.

We collected a total of 72 blink features from the IR sensor readings; 36 each from the SB and IB phases. These 72 blink features can be classified into five classes. (1) *IR Peaks:* This feature class measures the IR value returned at the apex of a blink, or when the users eye is closed. (2) *Rising IR:* This feature class measures the time it takes to close the eyelid starting from the time when the eye starts to close during the course of a blink. (3) *Falling IR:* This feature class measures the time it takes to go from a fully closed eyelid to back to an open eye during the course of a blink. (4) *IR peak interval:* This feature class measures the time between two blinks and captures how often people blink. (5) *IR floor:* This feature class measures the IR floor value representing the open eye

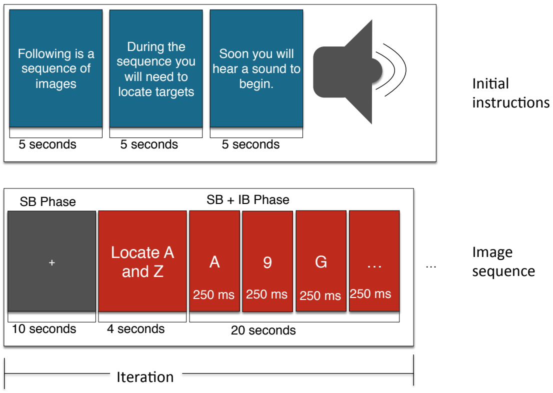

Figure 1. An annotated illustration of the image sequence shown to the user during the data collection process. Of its two stages, the first stage (in blue) informs the user what to expect during data collection. These slides are shown only once. The initial instructions are followed by an audio prompt which signals the start of the second stage – the actual data collection. The second stage (in red) shows the slides that are actually shown to the user. It always begins with a plus sign on the display. The slides from the second stage are repeated several times with different cue-cards.

state. Further, from each image sequence iteration we collected a total of 90 head-movement-related features.

The features can be classified into three categories. (1) *Gyroscope Peaks:* This feature class measures the direction of movement. (2) *Accelerometer Peaks:* This feature class measures the magnitude of movement observed for a user, particularly sudden head-movements (e.g., jerk). (3) *Movement time:* This feature class measures the time elapsed between various head-movements that were observed. For all three types of sensor readings (i.e., IR, gyroscope, and accelerometer), peaks were identified using a simple threshold, which was set to a single standard deviation above the mean.

Our goal in devising the original feature set was to select as many features as possible to ensure that in order to ensure that no pertinent features that can potentially influence user identification were missed. We therefore ranked the features based on the gain ratio, which is the probabilistic gain the feature adds toward correctly classifying a feature point and is measured as a rank in comparison to other features. Overall, we found a total of 96 *relevant features* from the original 162 features we started with. That is, we retained about 59.2% of the features we started with. In terms of the blink and movement features we retained 73.6% and 47.7% of the features, respectively. Please visit the following link (`http://hibou.cs.wpi.edu/~kven/iswc/features.pdf`) to see a list of the relevant features we used for this work.

Enrollment and Identification

Once the features are extracted for each user, the next step is to train a classifier for the group of users. We refer to this stage as *enrollment*. During enrollment, we capture sensor readings from each user in the group over a certain number of image sequence iterations. Each image sequence iteration produces a single feature point (feature vector), which is then used as a training sample for the classifier. Our method thus trains a classifier for *each user* by using that user's feature points as positive examples and all the feature points from other users as negative examples. This is the one-versus-all strategy used for multi-class classification. Using this approach, each new user requires the retraining of all models; however, we only need to collect samples from the new user because we can reuse the training data from the users who have already been enrolled. We examined several algorithms for the classifier including Random Forest, Support Vector Machine (SVM), and k-Nearest-Neighbor (kNN) and settled on using Random Forest (RF) with 100 trees. We chose RF because it copes well with the diverse, un-normalized features that have nonlinear relationships. In this work, we used 10 image sequence iterations to train the classifier. Once the classifier is trained, the next step is to use it to *identify* previously unobserved feature points from the users enrolled. A test feature point classified as positive for a given user is said to match that user's features; otherwise, the test feature vector is classified as negative.

EXPERIMENTAL SETUP

Dataset

In order to validate our approach, we needed data from the individuals wearing Google Glass. Therefore, we obtained approval from our university's institutional review board (IRB) and collected data from 20 volunteer users (15 males and 5 females with an average age of 22.25 years and SD of 3.7 years). The users were mostly drawn from the students on our campus. All data collection was conducted in a quiet space with adequate fluorescent lighting. We *did not* make an attempt to standardize the amount of lighting in the room where data was collected, beyond ensuring that the rooms were brightly lit. The data collection was performed both at our lab as well as at the users' homes. The users were informed of the purpose of our study and asked to sign a consent form approved by the IRB.

Evaluation Metrics

We performed a 10-fold cross-validation on the feature points to determine the efficacy of our approach. Given a set of test feature points being evaluated by the user-specific classifier, each test feature point from that specific user was labeled positive, while all the others were labeled negative, forming the ground-truth. We then used the aforementioned classifiers to classify all the test feature points, resulting in a positive or negative classification for each. Ideally, the classifier would classify only those test feature points belonging to the user as positive and everything else as negative. Given the classification results, we evaluated them based on the following metrics: false acceptance rate (FAR), false rejection rate (FRR), and balanced accuracy rate (BAC). The **false acceptance rate (FAR),** is the fraction of negatively labeled test feature points (the ground truth) that were misclassified as positive. The **false rejection rate (FRR),** is the fraction of positively labeled test feature points (the ground truth) that were misclassified as negative. The **balanced accuracy rate (BAC)**, is the sum of half of the *true acceptance rate (TAR)*, and half of the *true rejection rate (TRR)*. TAR is the fraction of positively labeled test feature points that were classified as positive, while TRR is the fraction of negatively labeled test feature points that were classified as negative. BAC is used instead of simple accuracy results to compensate for the sample imbalance (i.e., we have many more negative examples than positive ones) during the enrollment phase. *Even though we compute these metrics for every user in our dataset, we present summary statistics of these metrics over all users.* All our analysis was performed using WEKA [5].

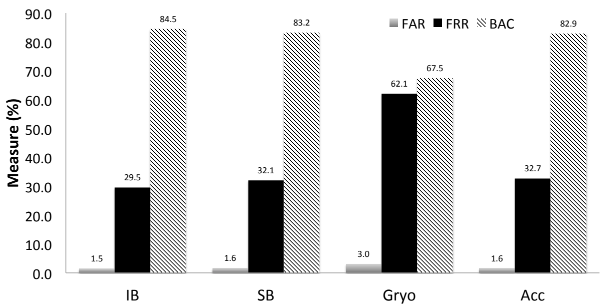

Figure 2. Performance of our approach w.r.t. the constituent feature groups. The blink features are broken down into constituent data collection phases, i.e., the SB phase and the IB phase. In the case of the movement features, the break-down is done based on the sensor types, i.e., accelerometer and gyroscope features. For each feature group we show the BAC, FRR and FAR metrics.

PERFORMANCE ANALYSIS

Overall Performance: To derive the overall performance, we ran the enrollment and identification processes a total three times for each of the 20 users and averaged the results. Our approach has a BAC of 94.4% with FRR of 11.3% and FAR of 0.5%. These results demonstrate that our approach is very effective in identifying users. The 0.5% FAR demonstrates that our approach makes very few errors in terms of identifying someone else as a particular user. However a ∼11% FRR indicates that the approach makes considerable mistakes in identifying a particular user as such. This makes the approach difficult to use, something we plan to address in future work.

Performance of Various Feature Groups: Given the overall performance, our next goal was to try to determine which feature groups within our model contributed the most to the identification process. Figure 2 shows the BAC, FRR and FAR for the various feature groups. The feature group breakdown is as follows. Note that the blink features collected during the IB phase, when the user is watching the image sequence, is the most important in identifying a user, while the gyroscope is the least important. Individually none of the features groups achieve over 85% accuracy, but together they improve the identification accuracy by over 10%. These results show that combining the various types of feature groups might be essential for robust identification.

These results demonstrate the *feasibility* of using blinking and head movement patterns for user identification without a self-identifying input from the user. However, there are still several concerns that need addressing. In addition to the aforementioned iteration duration, we may risk have a model prone to overfitting because of the use a very large feature set with only 20 users and three iterations. We suspect that by reducing the number of features to a handful, reducing the iteration duration, and increasing the number of users we can reduce the risk of overfitting and still obtain considerably accurate results. We are currently pursing this line of inquiry.

CONCLUSIONS

In this paper we presented a novel approach for identifying a user from a set of HMD users by utilizing their distinctive blinking and head-movement patterns. A preliminary analysis of our approach using data collected from 20 users demonstrates that our approach is over 94% accurate in identifying a user. In the immediate future we plan to extend this work in several directions including: (1) reducing the FRR and duration of each image sequence iteration from the current 34 seconds, (2) reducing the number of features in our approach, (3) evaluating the performance of our approach for different types of visual stimuli with varying degrees of content richness, and (4) studying the effectiveness of our approach for more diverse scenarios (e.g., effects of fatigue, effects of time elapsed between enrollment and identification on identification accuracy) and more diverse populations (e.g., older adults, children).

Acknowledgments

We would like to thank the anonymous reviewers and our shepherd for helping improve the quality of the paper.

REFERENCES

1. M. Abo-Zahhad, S.M. Ahmed, and S.N. Abbas. 2014. A new biometric modality for human authentication using eye blinking. In *Biomed. Engg. Conf. (CIBEC), 2014 Cairo International*. 174–177.
2. P. Baekgaard, M.K. Petersen, and J.E. Larsen. 2014. In the twinkling of an eye: Synchronization of EEG and eye tracking based on blink signatures. In *Cognitive Info. Proc. (CIP), 2014 4th Int. Workshop on*. 1–6.
3. Cory Cornelius, Ronald Peterson, Joseph Skinner, Ryan Halter, and David Kotz. 2014. A wearable system that knows who wears it. In *Int. Conf. on Mob. Sys., App., and Serv. (MobiSys)*. 55–67.
4. V. Genaro-Motti and K. Caine. 2014. Understanding the Wearability of Head-mounted Devices from a Human-centered Perspective. In *2014 ACM Int. Symp. on Wearable Comp. (ISWC '14)*. 83–86.
5. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA Data Mining Software: An Update. *SIGKDD Expl.* 11, 1 (2009).
6. G. Hogben. 2014. ENISA Briefing: Behavioral Biometrics. https://www.enisa.europa.eu/activities/risk-management/files/deliverables/behavioural- biometrics. (2014).
7. S. Ishimaru, K. Kunze, K. Kise, J. Weppner, A. Dengel, P. Lukowicz, and A. Bulling. 2004. In the blink of an eye: combining head motion and eye blink frequency for activity recognition with google glass. In *5th Augmented Human Int. Conf.* 15.
8. T. Nakano, Y. Yamamoto, K. Kitajo, T. Takahashi, and S. Kitazawa. 2009. Synchronization of spontaneous eyeblinks while viewing video stories. *Proceedings of the Royal Society B: Biological Sciences* 276, 1673 (2009), 36353644.
9. T. Westeyn and T. Starner. 2004. Recognizing song-based blink patterns: Applications for restricted and universal access. In *In Sixth IEEE Int. Conf. on Automatic Face and Gesture*. 717–722.