In the Blink of an Eye – Combining Head Motion and Eye Blink Frequency for Activity Recognition with Google Glass

Shoya Ishimaru, Kai Kunze, Koichi Kise
Osaka Prefecture University
Osaka, Japan
lastname@m.cs.osakafu-u.ac.jp, kise@cs.osakafu-u.ac.jp

Jens Weppner, Andreas Dengel, Paul Lukowicz
DFKI and University of Kaiserslautern
Kaiserslautern, Germany
firstname.lastname@dfki.de

Andreas Bulling
Max Planck Institute
Saarbruecken, Germany
bulling@mpi-inf.mpg.de

ABSTRACT
We demonstrate how information about eye blink frequency and head motion patterns derived from Google Glass sensors can be used to distinguish different types of high level activities. While it is well known that eye blink frequency is correlated with user activity, our aim is to show that (1) eye blink frequency data from an unobtrusive, commercial platform which is not a dedicated eye tracker is good enough to be useful and (2) that adding head motion patterns information significantly improves the recognition rates. The method is evaluated on a data set from an experiment containing five activity classes (reading, talking, watching TV, mathematical problem solving, and sawing) of eight participants showing 67% recognition accuracy for eye blinking only and 82% when extended with head motion patterns.

Author Keywords
Activity Recognition; Blink Frequency; Infrared Proximity Sensor; IMU; Head Mounted Sensor; Google Glass

ACM Classification Keywords
I.5.4 PATTERN RECOGNITION Applications: Signal processing

INTRODUCTION
For a long time sensors mounted on the user's head were seen as too obtrusive for activity recognition in every day life scenarios (as opposed to for example industrial applications where sensors could easily be integrated in e.g. helmets). The Google Glass platform (and a score of emerging similar devices) has clearly undermined this assumption. It has been designed for all day use in every day situations and, over the last year, has been used in this way by thousands of people.

Google Glass has four sensors that could potentially be used for activity recognition: a camera, a microphone, an inertial measurement unit (IMU), and an infrared proximity sensor facing towards the users' eye that can be used for blink detection. In this paper we focus on the later two (IMU and blink detection) which we argue to be most characteristic to the Google Glass (and similar) platform. For the microphone the position on the head makes little difference (except possibly for recognizing the user speaking). There is also a lot of existing work on head mounted cameras. On the other hand combining eye blink frequency and head motion patterns for activity recognition in every day scenarios has so far not been studied in much detail.

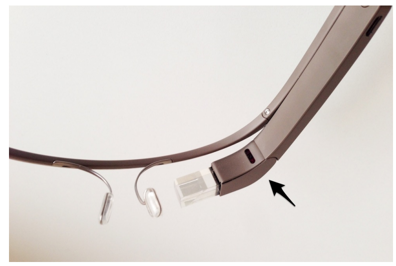
Figure 1. Infrared proximity sensor built into Google Glass

Paper Contributions
From the above considerations this paper investigates how a combination of eye blink detection and head motion pattern analysis can be used to recognize complex high level human activities. Overall the aim is not to present a "ready for use" method for the recognition of specific applications. Instead, it is to explore the limits of how much information can be extracted from a very simple blink detector provided by the Google Glass platform and how important additional information about head motion is. Thus on a practical level one contribution is to discuss the blink recognition with Google Glass. While Google provides an undocumented API to detect eye blinks as a control method for the system, we argue that a detailed description and evaluation of long term statistical analysis of blinking frequencies provided in this paper is a valuable contribution that could also benefit researchers using other similar setups.

On a conceptual level the main contribution is to show that the proposed methodology has the potential to contribute to the recognition of activities that are difficult to distinguish using other, unobtrusive wearable sensors. Specifically, we consider reading, talking (having a conversation), watching

AH '14, March 07 - 09 2014, Kobe, Japan.
Copyright 2014 ACM 978-1-4503-2761-9/14/03...$15.00.
http://dx.doi.org/10.1145/2582051.2582066

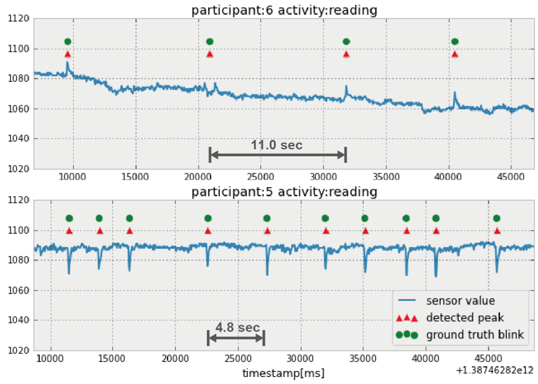

Figure 2. Proximity sensor value and ground truth of 2 participants.

a video, solving a mathematical problem and sawing. Distinguishing between these activities involves not only recognizing physical actions (that can easily be captured using for example on body motion sensors) but also a cognitive component which is what we hypothesize eye blinking frequency and head motion correlate with.

We evaluate our method on a data set containing eight participants demonstrating an average classification accuracy of 67% using blink features only and 82% using blink and motion features.

Related Work

There is a large corpus of work to recognize human activities. A variety of physical activities can be recognized using body-mounted sensors [5]. On the other hand, some researchers focus on our cognitive activities. Bentivoglio et al. have studied the relation between sitting activities and blink patterns [3]. They described that the blink rate changes when participants were reading, talking and resting. Acosta et al. have presented that working with computers causes a reduction of blink [1]. Haak et al. have described that emotion, especially stress, effects blink frequency [9]. Therefore, blink pattern should be one of the important features to recognize our activities. Some researchers have applied an image processing method [6] and an eye tracking approach [4] to detect blinks.

As far as we know, we are the first to use a simple proximity sensor embedded in a commercial wearable computing system for activity recognition and to combine it with head motion patterns.

APPROACH

We believe that blink patterns can give a lot of insights about the user's mental state (drowsiness etc.) and the user's activity. To show this we use an infrared proximity sensor on Google Glass (see Figure 1). It monitors the distance between the Google Glass and the eye. Figure 2 shows the raw values of the sensor. While the main function of this sensor is to detect if the user wears the device, when the user blinks, a peak value appears due to the eye lid and eyelashes movement. Our algorithm is based on two stages. The first stage is the pre-processing stage of the raw sensor signal. The pre-processing stage extracts the time of blinks. We validate the pre-processing results with ground truth blink information.

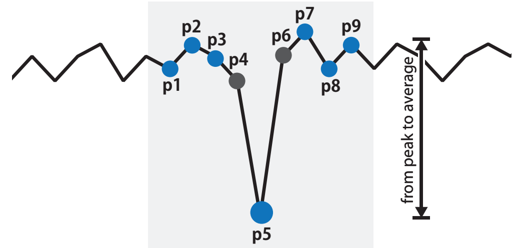

Figure 3. Blink detection by calculating peak value.

Secondly, the main part of our algorithm calculates features based on the detected blinks. Getting raw data of infrared proximity sensor on Google Glass is not provided in an official way. We rooted (customized) our Glass on the basis of Glass hacking tutorial [7] and installed our own logging application [8] for the experiment.

Blink detection

During pre-processing blinks are detected based on the raw infrared proximity sensor signal. We move a sliding window on the sensor data stream and monitor whether the center point of each window is a peak or not according to the following definition. We calculate the distance from one sensor value of the center point in the window (p_5 in Figure 3) to the average value of other points (p_1, p_2, p_3, p_7, p_8 and p_9). The preceding and subsequent points of the center (p_4 and p_6) are excluded from the average calculation because their sensor values are often affected by the center point. If the distance is larger than a threshold ranging from 3.0 - 7.0 we define the center point as a blink. Because the shape of the face and eye location vary, the best threshold for the peak detection varies for each user. Figure 2 with the same scale for each sub-graphic also demonstrates different signal variations for different users. We calculate the best threshold (in 0.1 steps ranging from 3.0 to 7.0) by evaluating the accuracy based on the ground truth information. This approach can be applied only in off-line evaluation. In on-line usage, we need a few seconds for calibration before detection. During the calibration term, Glass urges the user to blink as matching some timing. We get sensor values and actual blink timing from calibration and evaluate the best threshold.

Blink frequency based activity recognition

As an output of our pre-processing step we extract the timestamps of blinks and compute a three-dimensional feature vector. One is the mean blink frequency which describes the number of blinks during a period divided by the length of a period. Two other features are based on the distribution of blinks. Graphically, this can be understood as the histogram of the blink frequency. Figure 5 shows five histograms with a period of 5 minutes. The x-axis describes the mean blink frequency (0.0 - 1.0 Hz) and the y-axis describes the blink counts of each frequency. The number of specified bins per histogram is 20 having a resolution of 0.05 Hz. The frequency value is calculated as inverse value of the interval between two blinks. The second and third features are defined as the x-center of mass and the y-center of mass of the histogram.

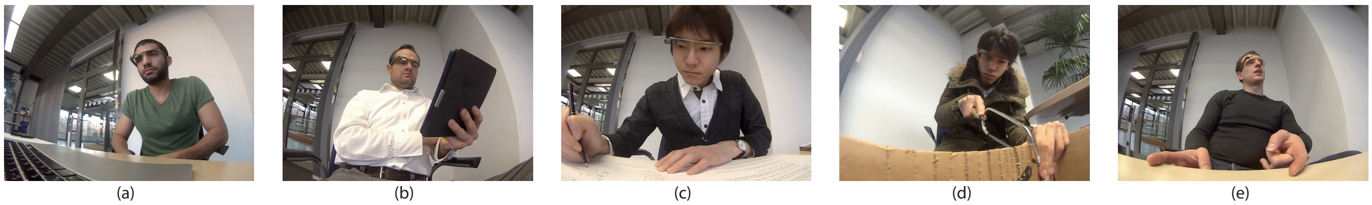

Figure 4. Video based ground truth image excerpts of the experiment scenes containing watching (a), reading (b), solving (c), sawing (d) and talking (e).

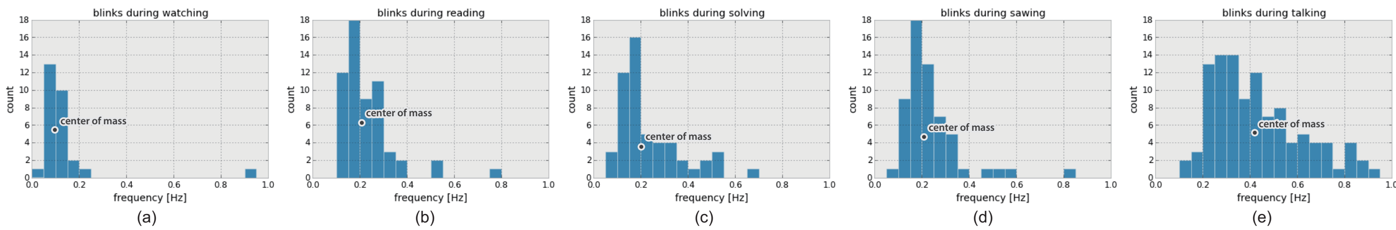

Figure 5. Overview of blink frequency during 5 minutes activity, watching (a), reading (b), solving (c), sawing (d) and talking (e).

Table 1. Dataset overview from ground truth

Participant	1	2	3	4	5	6	7	8	Avg.
Total blink counts [1]	230	161	420	313	381	309	207	414	**304**
Min frequency (Hz)	0.02	0.02	0.03	0.06	0.02	0.04	0.02	0.02	**0.03**
Max frequency (Hz)	0.96	0.96	0.99	0.99	0.99	0.98	0.99	0.99	**0.98**

[1] The total blink counts through all activities for each participant

Combination of blink frequency and head motion

The advantage to use Google Glass is that we can easily get and combine several sensor signals for our activity recognition task. The user's head motion varies for different activities. We calculate the degree of head motion by calculating the averaged variance of the three-dimensional accelerometer sensor signal which defines our simple motion based feature. In the approach combining blink patterns and head motion we use the following four features: variance value of accelerometer, the mean value of blink frequency and the x-center and y-center of mass value of blink frequency histogram. We combine these features and compare the impact on activity recognition accuracy.

EVALUATION AND RESULTS

We recruited eight participants to perform five activities each lasting five minutes while wearing the Google Glass. All of the participants were male. Five of them had unaided vision and three (2, 3, 4) were using contact lenses. We defined the activities as watching a movie on a Laptop, reading a book on an ebook reader, solving mathematical problems on paper (entrance examination for graduate school), sawing a cardboard and talking with another person. We intended solving as an example of mental tasks and sawing as a physical task. The location and light condition was fixed for all experiment participants. The display of Google Glass was always turned off and didn't attract the subject's attention during the experiment. We collected values of the infrared proximity sensor and the accelerometer. Each activity was recorded separately. Feature extraction and classification was applied to the data containing a single activity. We also recorded the experiment scene video to get ground truth. Figure 4 shows the ground truth information of five different users performing five different activities. After the experiments we used the video as the ground truth information for activity classes and primarily for labeling every participant's blink timing by using Labeling Tool [2] for pre-processing step evaluation.

Table 2. Pre-processing blink detection results

Participant	1	2	3	4	5	6	7	8	Avg.
Accuracy (%)	96	85	89	99	97	93	94	89	**93**
Precision (%)	92	48	76	98	87	80	71	86	**80**
Recall (%)	64	71	72	98	89	86	72	74	**78**

Table 3. Activity recognition classification results

Participant	1	2	3	4	5	6	7	8	Avg.
Method 1 (%) [1]	70	52	82	76	70	54	69	64	**67**
Method 2 (%) [2]	75	56	66	83	57	56	58	50	**63**
Method 3 (%) [3]	92	81	87	91	82	74	74	74	**82**

[1] Blink frequency based recognition.
[2] Head motion (accelerometer) based recognition.
[3] Combination of blink frequency and head motion by fusing infrared proximity sensor data and accelerometer data.

At the recognition part we defined the sliding window size as 60 seconds with a step size of 10 seconds and calculated the previously defined features for each window. The window size should be longer than max interval in dataset. The longest blink interval through all participants was 50 seconds (see Table 1 for details). We trained a user dependent J48 decision tree classifier and evaluated the classification accuracy by confusion matrices based on 10-fold cross validation.

Blink detection

Figure 5 exemplarily shows for one participant five different histograms based on the blink frequency distribution during five minutes for each activity. We evaluated the blink detection according to our ground truth and achieved an average accuracy over all participants of 93% ranging from 85% to 99% (see Table 2 for details). Each participant's blink detection accuracies are based on the average value of 5 activities.

Activity recognition

Solely based on blink frequency features and an experimental complexity of eight participants and five activity classes we

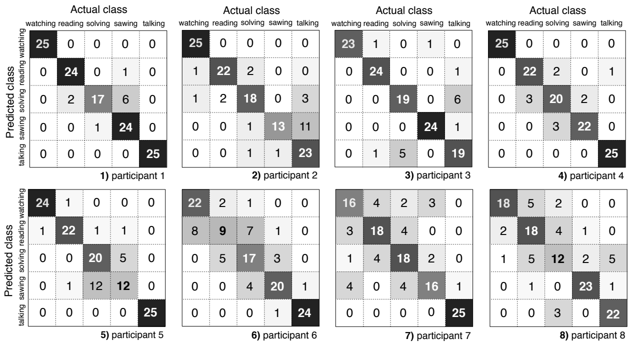

Figure 6. Confusion matrices of all participants.

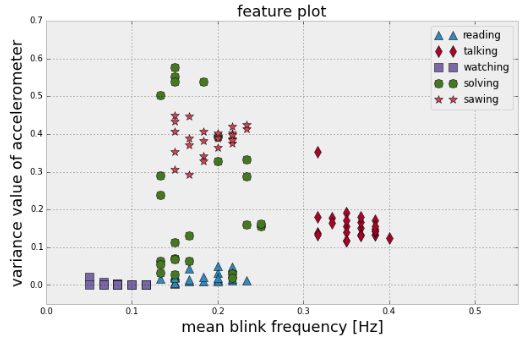

Figure 7. Feature representation of one person and five different activity classes are shown. Each dot represents a data segment of 60 seconds.

achieved an average classification accuracy of 67% (see Table 3 for an overview of all participants) individually ranging from 52% to 82%. Solely motion feature based recognition underperformed with 63% classification accuracy. When we combine the blink frequency based features with the motion based feature we achieve an average classification accuracy of 82% (increased by 15% compared to blink frequency based recognition). Figure 6 shows the individual confusion matrix results of eight experiment participants. These confusion matrices show correctly classified instances on the diagonal and wrongly classified instances in other areas.

Figure 7 shows one participant's feature plot. Talking and watching is easily distinguished by other activities. Yet it is difficult to classify sawing, reading and solving by only blink patterns. Head motion feature helps to distinguish especially those classes. Conversely, reading and watching can not be distinguished easily only by head motion. The dispersion of head motion during solving is larger than other activities because solving contains 2 statuses, concentrating to write the answer and looking at the assignment on another paper.

The training duration per class and per person was only five minutes long. In future the input of the correct activity might be given during daily usage of Google Glass learning constantly from the user's activities and improving the classification constantly. We evaluated ten minutes of recordings of six participants (1, 4, 5, 6, 7 and 8) again. The classification based on blink frequency improved by 7% an in combination with the motion feature improved by 9% compared to the five minute long recording.

CONCLUSION AND FUTURE WORK

We have shown how the infrared proximity sensor from the standard Google Glass can be used to acquire user eye blink statistics and how such statistics can be combined with head motion pattern information for the recognition of complex high level activities. The overall accuracy of 82% for the selected five activity classes may still not be enough for many practical applications. However, it clearly indicates that these sensors contain valuable information which could complement other, more widely used sensors such as as sound, vision, and motion.

A possible improvement of the system itself includes a selection of valid infrared proximity sensor signal segments by filtering out parts with significant movement of the user. For further improvement in this method future work will focus on a better understanding of the optimal test segment size and of the changes in blink frequency in different conditions (depending on air humidity etc).

REFERENCES

1. Acosta, M. C., Gallar, J., and Belmonte, C. The influence of eye solutions on blinking and ocular comfort at rest and during work at video display terminals. *Experimental eye research 68*, 6 (1999), 663–669.
2. Bannach, D., Kunze, K., Weppner, J., and Lukowicz, P. Integrated tool chain for recording and handling large, multimodal context recognition data sets. In *Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing-Adjunct*, ACM (2010), 357–358.
3. Bentivoglio, A. R., Bressman, S. B., Cassetta, E., Carretta, D., Tonali, P., and Albanese, A. Analysis of blink rate patterns in normal subjects. *Movement Disorders 12*, 6 (1997), 1028–1034.
4. Bulling, A., Ward, J. A., Gellersen, H., and Tröster, G. Eye movement analysis for activity recognition. In *Proceedings of the 11th international conference on Ubiquitous computing*, ACM (2009), 41–50.
5. Casale, P., Pujol, O., and Radeva, P. Human activity recognition from accelerometer data using a wearable device. In *Pattern Recognition and Image Analysis*. Springer, 2011, 289–296.
6. Chau, M., and Betke, M. Real time eye tracking and blink detection with usb cameras. Tech. rep., Boston University Computer Science Department, 2005.
7. Voiding Your Warranty: Hacking Glass. `https://developers.google.com/events/io/sessions/332704837`.
8. Mrk1869/GlassLogger - GitHub. `https://github.com/Mrk1869/GlassLogger`.
9. Haak, M., Bos, S., Panic, S., and Rothkrantz, L. Detecting stress using eye blinks and brain activity from eeg signals. *Proceeding of the 1st Driver Car Interaction and Interface (DCII 2008)* (2009).