Unintrusive Eating Recognition using Google Glass

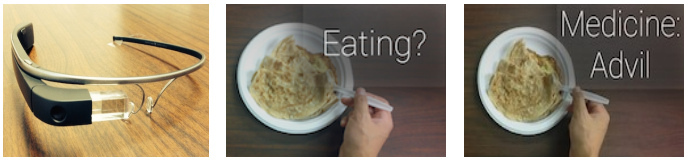
Shah Atiqur Rahman, Christopher Merck, Yuxiao Huang, Samantha Kleinberg
Stevens Institute of Technology
Hoboken, NJ
{srahman1,cmerck,yhuang23,samantha.kleinberg}@stevens.edu

***Abstract*—Activity recognition has many health applications, from helping individuals track meals and exercise to providing treatment reminders to people with chronic illness and improving closed-loop control of diabetes. While eating is one of the most fundamental health-related activities, it has proven difficult to recognize accurately and unobtrusively. Body-worn and environmental sensors lack the needed specificity, while acoustic and accelerometer sensors worn around the neck may be intrusive and uncomfortable. We propose a new approach to identifying eating based on head movement data from Google Glass. We develop the Glass Eating and Motion (GLEAM) dataset using sensor data collected from 38 participants conducting a series of activities including eating. We demonstrate that head movement data are sufficient to allow recognition of eating with high precision and minimal impact on privacy and comfort.**

I. Introduction

Chronic diseases such as obesity and diabetes affect an increasing portion of the population and require new patient-centered management strategies as patients and their caregivers have the primary day-to-day management responsibility. Key problems include providing patient-centered decision support (e.g. adjusting insulin doses for people with diabetes), improving medication adherence, and creating logs to be reviewed with clinicians. Many of these needs center on eating: nutrition is critical to managing blood glucose, many medications are taken with food or on an empty stomach, and nutrition logs for obesity and other diseases have low adherence and accuracy.

Activity recognition has previously been used for applications such as predicting falls in the elderly using mobile phone accelerometers, identifying running to track exercise, and understanding movement patterns through a home. While eating has been less studied, it is critical for three areas of health. Eating detection will enable personal informatics to automate meal logging and give feedback to users. Second, it can improve the interaction between individuals and their environment [4]. For example, an application can detect a user is eating and silence her phone to avoid interruption. Finally, there is a strong connection between eating and health. People with chronic disease, a rapidly growing portion of the population, may particularly benefit due to the importance of self-management in the treatment of these diseases. Eating detection can support healthy individuals and those with chronic and acute disease by logging and rewarding activity (through gamification mechanisms), providing contextual medication reminders (as many medications must be taken with food or on an empty stomach), or prompting a person with diabetes to measure their blood glucose. Systems that provide frequent feedback can improve treatment adherence, but this requires accurate systems that do not intrude on daily life.



(a) (b) (c)

Fig. 1: Device and sample application: (a) Google Glass, (b) action recognition and (c) reminder in response to action.

We propose that continuously collected data from unintrusive head-mounted sensors can be used to recognize eating and other activities. Our pilot Glass Eating and Motion (GLEAM) dataset shows that the sensors in Google Glass (e.g. accelerometer, gyroscope) are sensitive enough for this purpose. With a combination of machine learning methods, we achieve high accuracy for eating detection. While we focus on the use of Glass because of its flexibility for prototyping and the possibility of sending feedback to users as shown in figure 1, in the future other head-mounted sensors (e.g. earbuds) may be developed when visual feedback is not needed.

II. Related Work

Much prior work has been done on activity recognition, in particular on recognizing locomotion using data from accelerometers in cellphones or body-worn sensors [11]. Less work has been done on detecting eating, with the main approaches using audio or image data, environmental sensors such as RFID, or sensors placed around the throat. The use of acoustic sensors (to detect chewing sounds) ([9], [2]) or cameras [12] raises major privacy concerns and may face challenges due to lighting or background noise, while environmental sensors [3] have limited generalizability as they depend on a controlled and tagged environment. Finally, the use of sensors to detect swallowing [1] requires these devices to be placed around the throat, which can be uncomfortable.

Ishimaru et. al. have investigated head movement in conjunction with blink detection using Google Glass [6], but did not apply the technique to eating detection. They also tested electrooculography glasses for discriminating several activities including eating [7], but this study was limited by a small sample size (2 participants).

III. Data Collection

To test our hypothesis that head movement can be used to recognize eating, we developed the GLEAM dataset, using data collected from 38 participants wearing Glass while conducting

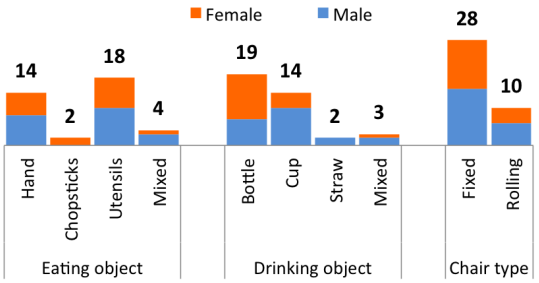


Fig. 2: Demographic information.

a series of activities in a controlled environment. Activity times were annotated by two of the researchers during data collection. Participants provided informed consent and the protocol was approved by the Stevens IRB.[1]

A. Procedure

Our main goal is detecting eating, but this is a small portion of a person's daily activities. Thus, our protocol involved 2 hours of data collection spanning eating, brief walks and other activities. To ensure that the eating data were representative of usual behavior, participants brought their own meals (usually lunch) to our lab space. Meals included pasta, bagels, pizza, sandwiches, sushi, yogurt, salad, and nuts.

Participants talked with the researcher, ate a meal in two parts (to yield two onset times) with a 5 minute break in-between, walked down and up stairs and across a hallway, drank a beverage, and performed other activities of their choice (e.g. reading, working on a computer) until 2 hours had elapsed. The primary activity category was "other." Not all participants performed the activities in the same order and they were permitted to perform multiple activities simultaneously.

Participants wore Glass for the duration of data collection but did not interact with the device to avoid biasing the data toward gestures like swipes and taps. Instead, Glass's sensors recorded movement and a researcher annotated activity start and end times on a separate device.

B. Sensors

Data was collected from participants wearing Google Glass, which has a similar form factor to glasses but with a display and no lenses. Along one arm (on the right side of the head), as shown in figure 1a, Glass contains several sensors and computing hardware similar to that of a cellphone. The sensors include: accelerometer, gyroscope, magnetometer, and light sensor. The Glass API allows reading these sensors directly and provides processed values for gravity, linear acceleration, and a rotation vector (a quaternion representing the device's orientation). We developed Glassware that collects data from all sensors and processed sensor values except light with a median sampling period of 395 ms. The sampling rate was chosen as a trade-off between time-granularity and battery life.

[1]Data are available at: http://www.skleinberg.org/data.html

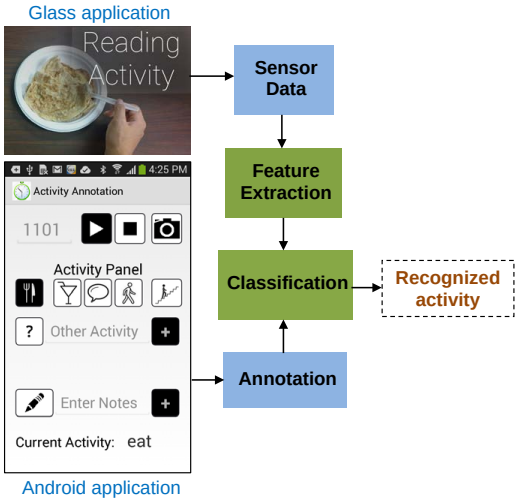


Fig. 3: Overview of activity recognition.

	A1	A2	#	p-value
Onset	0.75	1.07	75	0.6944
Offset	0.79	1.12	75	0.2955
Eat	0.56	0.76	62	0.2612
Drink	1.14	1.32	22	0.7473
Walk	0.50	1.50	22	0.3577
Talk	0.91	1.27	44	0.2147

TABLE I: Mean annotation errors (in seconds) for two annotators (A1, A2). P-values are based on a paired t-test.

C. Participants

Data were collected from 21 male and 17 female participants, aged 18-21 (median 20), recruited from the University and surrounding area. We excluded individuals with prior Lasik surgery (based on Google's Glass guidance) and those with difficulty chewing or swallowing. All participants completed the full 2 hours of data collection.

Implements used for eating, drinking, and sitting are shown in figure 2. Additionally, 3 participants wore Glass atop their own prescription glasses. Pictures of food were taken prior to eating, after half the meal was completed, and at the end to enable analysis of accuracy by food type and meal size.

D. Annotation

Researchers observing the participants annotated the start and end times of all activities using an Android app we developed for the Samsung Galaxy cellphone, as shown in figure 3. The clocks of Glass and the cellphone were synchronized. Activity onsets were defined as follows. Eating begins when food is first placed in the mouth, talking begins when a word is spoken, drinking begins when a beverage or straw is placed in the mouth, and walking or navigating stairs begins with the first footfall. If a participant performed several activities at once, only the dominant activity was annotated. Thus the end of one activity is the beginning of another.

To ensure the quality of annotation, we conducted preliminary tests of inter-rater reliability. Two research group

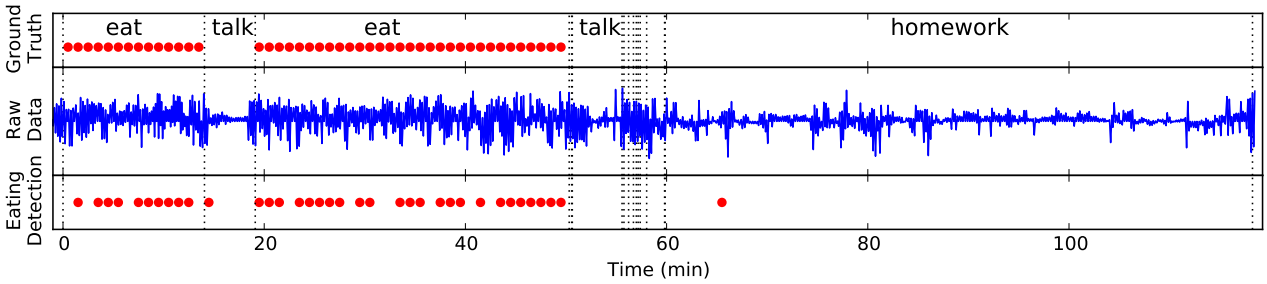


Fig. 4: Activity annotations, accelerometer x-axis data, and eating detected by LOWO RF classifier for one participant.

members were video recorded as they ate, drank, walked and talked, with a total of 75 activities (150 combined onsets and offsets). The videos were then independently annotated by two researchers. The mean absolute error in seconds between each annotator and video (treated as ground truth) for each of the four main activities are shown in table I. Across all activities, the mean error is 0.77 and 1.10 seconds for annotators 1 and 2 respectively. Onset errors are lower than for offset, but using a paired t-test, no differences were statistically significant ($p > 0.2147$ in all cases).

IV. Eating Recognition

The key steps of the activity recognition process are extracting features from the sensor data and training the classifier.

A. Feature Extraction

The raw sensor data was divided into minute-long non-overlapping windows, with the window length chosen to be long enough to capture periodic head movements but short enough for granular activity onset detection. To explore the effect of window boundaries on classification performance, we generated features based on 6 different window offsets (0 through 50 seconds with 10 second steps).

From each window we generate a 180-element vector consisting of features proven useful for activity recognition ([2], [13]), specifically 5 statistical (mean, intensity, variance, skewness, kurtosis and crest-factor), 3 spectral (spectral centroid, flux, and roll-off), and 2 temporal (delta and delta-delta coefficient [8]) features, for each of 6 sensors and 3 axes. Each feature vector is labeled with the most frequent activity in the window. An example of raw data along with ground-truth and predicted class labels is shown in figure 4.

B. Cross-Validation

Due to heterogeneity between participants, we expect classification performance will be improved by training on some individual data. To evaluate the relationship between the amount of training data for a given participant and the classifier performance, we tested three cross-validation approaches which use progressively more training data: *leave one participant out* (LOPO), *leave half participant out* (LHPO), and *leave one window out* (LOWO).

LOPO has 38 cross-validation folds, each created by training a classifier on data from 37 participants and testing it on the held out one. In LHPO each participant's data is divided into two parts by splitting each labeled activity in half. In each of 76 folds, a classifier is trained on data from 37 participants and half of the data from the remaining participant. Finally in LOWO, all data is used for training except for a single 1-minute window, which is then used for evaluating the classifier.

C. Classification

Focusing on the classification between eating and other (primarily reading and homework), we evaluated Gaussian Naive Bayes (NB) and k-nearest neighbor (kNN) with $k = 1$ due to their prevalence in related work. We also evaluated C4.5 decision tree and Random Forest (RF) with 10 trees due to their robustness to non-informative features. As the performance of NB and kNN are negatively impacted by the presence of non-informative features, we trained these classifiers only on the 10 most informative features, selected using an Information Gain criterion. All classification and feature selection was performed using implementations in WEKA [5] with default parameters.

D. Performance Metric

The appropriate choice of performance metric is data and application dependent. Our data has a strong class imbalance (only 11.57% of samples are of eating, while 72.43% are homework/free time), so an F_β score, where β controls relative importance of recall and precision, is most appropriate. Our envisioned applications (e.g. automated meal logging, medication reminders) are sensitive to false positives, which would result in annoying spurious alerts and logged events, and therefore we particularly value precision. Furthermore, because our annotations indicate start and end times of entire meals, false negatives may actually be due to pauses between bites or courses. Therefore we report the $F_{0.5}$ score, which weights precision more strongly than recall, as well as the Area Under Curve (AUC) for comparison with related work.

V. Experimental Results

The $F_{0.5}$ score for each classifier and cross-validation approach is given in table II. Corresponding confusion matrices are shown in table III. Tree-based classification methods performed well due to implicit feature selection, and more training data consistently improved performance, with the LOWO RF classifier performing best. Window offset was not found to significantly impact performance ($p = 0.92$ using one-way ANOVA of $F_{0.5}$ score for LOWO RF).

Results varied between participants, as shown in figure 5. No eating was detected in 9 participants, including two who

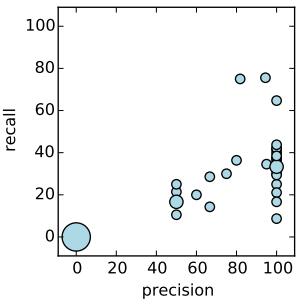


Fig. 5: Performance of LOWO RF classifier for eating class across 38 participants. The marker areas are proportional to number of participants with identical performance.

frequently adjusted their prescription glasses, one who moved abruptly in a rolling chair, and one who ate only a small meal of pudding and cookies. Shorter meals (less than 15 minutes) were less likely to be recognized. However, for 11 participants 100% precision was achieved. The results are robust with respect to the objects used for eating, with no significant difference in LOWO RF $F_{0.5}$ score ($p = 0.80$ using one-way ANOVA). Similarly, sitting on a rolling versus fixed chair had no significant effect ($p = 0.52$ using unpaired t-test).

The performance achieved here is better than that reported by other approaches to eating detection, showing the value of using head movement for this purpose. For example, Logan et al. [10] used more than 900 environmental sensors tagged with objects and achieved AUC of 0.587 for eating recognition whereas our AUC is 0.92 for RF. Finally, the more intrusive audio visual data of Liu et al. [9] suffered from a high false positive rate: 13.07% of "other" class detected as eating in contrast to only 1.79% in our study.

The performance of our LOWO RF classifier relies on user-specific training data. Although in our study participants did not interact with Glass, future studies may evaluate an active learning approach where the Glass app prompts users for feedback to gather the required training data so that eating detection may be iteratively improved.

VI. Conclusion

While eating recognition has lagged behind that of other activities, newly developed head-mounted sensors have made this feasible with high accuracy in an unobtrusive and privacy-preserving manner. We propose a new approach to recognizing eating from head movement that can facilitate automated nutrition logs, smart medication reminders, and individualized chronic disease management. The data collected are a realistic sample of eating and drinking as they contain significant natural variation, such as participants doing multiple activities at once and eating their own foods with their chosen utensils. Even without developing algorithms specifically for this purpose, our study achieved a higher $F_{0.5}$ (67.55%) for eating recognition than the best reported methods, and we demonstrated robustness to utensils used. Future work may focus on active learning to improve performance over time, determining activity onset time and duration, and using the meal photographs to estimate meal size and type.

Classifier	Metric	LOPO	LHPO	LOWO
k-NN	$F_{0.5}$	31.27%	33.79%	37.06%
	AUC	0.638	0.644	0.652
NB	$F_{0.5}$	19.71%	20.16%	19.19%
	AUC	0.786	0.804	0.779
C4.5	$F_{0.5}$	41.95%	50.0%	53.53%
	AUC	0.639	0.611	0.677
RF	$F_{0.5}$	49.73%	58.77%	67.55%
	AUC	0.858	0.884	0.922

TABLE II: $F_{0.5}$ score and AUC for different classifiers and different cross validation approaches.

	k-NN		NB		C4.5		RF	
	E	O	E	O	E	O	E	O
E	213	325	495	43	273	265	244	294
O	371	3708	2595	1484	230	3849	73	4006

TABLE III: Confusion matrices for eat (E) and other (O) for LOWO. Rows are actual and columns classified label.

Acknowledgments

We thank Jason Gardella and Mark Mirtchouk for their assistance. Data collection was supported in part by the Center for Healthcare Innovation at Stevens. SK was supported in part by NSF Award #1347119.

References

[1] O. Amft and G. Troster, "On-Body Sensing Solutions for Automatic Dietary Monitoring," *IEEE Pervasive Computing*, vol. 8, no. 2, 2009.

[2] Y. Bi, W. Xu, N. Guan, Y. Wei, and W. Yi, "Pervasive eating habits monitoring and recognition through a wearable acoustic sensor," in *Pervasive Health*, 2014.

[3] M. Buettner, R. Prasad, M. Philipose, and D. Wetherall, "Recognizing Daily Activities with RFID-based Sensors," in *UbiComp*, 2009.

[4] J. García-Rodríguez and J. M. García-Chamizo, "Surveillance and Human-computer Interaction Applications of Self-growing Models," *Appl. Soft Comput.*, vol. 11, no. 7, 2011.

[5] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, 2009.

[6] S. Ishimaru, K. Kunze, and K. Kise et al., "In the blink of an eye: Combining head motion and eye blink frequency for activity recognition with google glass," in *Augmented Human*, 2014, pp. 15:1–15:4.

[7] S. Ishimaru, Y. Uema, K. Kunze, K. Kise, K. Tanaka, and M. Inami, "Smarter eyewear: using commercial eog glasses for activity recognition," in *UbiComp Adjunct*, 2014, pp. 239–242.

[8] D.-S. Kim, S.-Y. Lee, and R. Kil, "Auditory processing of speech signals for robust speech recognition in real-world noisy environments," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 1, 1999.

[9] J. Liu, E. Johns, L. Atallah, C. Pettitt, B. Lo, G. Frost, and G.-Z. Yang, "An intelligent food-intake monitoring system using wearable sensors," in *Proc. BSN*, 2012.

[10] B. Logan, J. Healey, M. Philipose, E. M. Tapia, and S. Intille, "A long-term evaluation of sensing modalities for activity recognition," in *UbiComp*, 2007.

[11] T. Plötz, N. Y. Hammerla, and P. Olivier, "Feature Learning for Activity Recognition in Ubiquitous Computing," in *IJCAI*, 2011.

[12] E. Thomaz, A. Parnami, J. Bidwell, I. Essa, and G. D. Abowd, "Technological Approaches for Addressing Privacy Concerns when Recognizing Eating Behaviors with Wearable Cameras," in *UbiComp*, 2013.

[13] K. Yatani and K. N. Truong, "Bodyscope: a wearable acoustic sensor for activity recognition," in *UbiComp*, 2012, pp. 341–350.