When Generalized Eating Detection Machine Learning Models Fail in the Field

Shibo Zhang
Northwestern University
Chicago, IL 60611, USA
shibozhang2015@u.northwestern.edu

Rawan Alharbi
Northwestern University
Chicago, IL 60611, USA
rawan.alharbi@northwestern.edu

Matthew Nicholson
Northwestern University
Chicago, IL 60611, USA
matthewnicholson2019@u.northwestern.edu

Nabil Alshurafa
Northwestern University
Chicago, IL 60611, USA
nabil@northwestern.edu

UbiComp/ISWC'17 Adjunct , September 11–15, 2017, Maui, HI, USA.
 ISBN 978-1-4503-5190-4/17/09...$15.00.
DOI: https://doi.org/10.1145/3123024.3124409

Abstract
Problematic eating behaviors are a major cause of obesity. To improve our understanding of these eating behaviors, we need to be able to first reliably detect them. In this paper we use a wrist-worn sensor to test a generalized machine learning models' reliability in detecting eating episodes through data processing. We process data from a 6-axis inertial sensor. Since most eating episodes do not occur while moving, we filter out periods of physical activity, and then use an advanced motif-based time-point fusion technique to detect feeding gestures. We also cluster each of the false alarms into four categories in an effort to identify the main behaviors that confound feeding gesture detection. We tested our system on eight participants performing various activities in the wild while wearing a sensing suite: a neck- and a wrist-worn sensor, along with a wearable video camera continuously recording to capture ground truth. Trained annotators further validated the algorithms by identifying feeding gestures, and categorized the false alarms. All eating episodes were detected; however, many false alarms were also detected, yielding a 61% average F-measure in detecting feeding gestures. This result shows clear challenges in characterizing eating episodes by using a single inertial-based wrist-worn sensor.

Author Keywords
Wrist-worn sensors; wearables; hand-to-mouth gestures;

in-the-field test; overeating; inertial sensors; motif-based segmentation; K-Spectral Centroid clustering; fusion; classification; feeding gesture.

ACM Classification Keywords
H.1.2 [User/Machine Systems]: Human Factors

Introduction
More than two-thirds of adults are considered to be obese or overweight [9]. Obesity costs more than 147 billion dollars per year in the US [13]. Studies on obesity and diet have been main focuses in the behavioral sciences and the health care community for many years. One traditional method of recording dietary intake relies on self-reporting data, which may result in high burden on participants and high bias over long periods of time. With the development of wearable sensors, pervasive computing, and passive sensing data analytics, various automated dietary recording systems have been developed for eating behavior recognition and analysis in recent decades [7, 14, 6, 3, 10, 4]. Developing such systems includes embedded system design as well as signal processing, human activity recognition, and behavioral analysis.

There are many different wearable sensors that exist on the market today. However, given the ubiquitous nature of smart watches, people are most willing to wear a wrist-worn sensor compared to sensors worn on other parts of the body [1]. Feeding detection is one application of sporadical activity spotting, which is a difficult and open problem within the activity recognition community [7].

The challenges of feeding gesture detection include reliable data collection (proper mapping of data to labels), inaccurate gesture recognition due to interclass similarity of confounding gestures (due to the high volume of possible confounding gestures), and high intraclass variability of human feeding gestures (due to variability across culture and individual habits). Simply dividing the data into feeding gestures and non-feeding gestures (NULL class) places several gestures that are similar to feeding within the NULL class. Being able to explain which activities within the NULL class are confounding behaviors will help us improve our machine learning models for detection and characterization of eating behaviors.

In order to build an effective activity recognition model that can generalize well in free-living populations, structured and unstructured eating activities, along with confounding activities, are induced in the lab. Increasingly, studies have incorporated in field data into their model development, resulting in more realistic gestures and human behaviors.

However, the majority of prior work in this field aims to build models from in-lab data [3, 4, 12, 8]. While it is known that personalized models outperform generalized models in predictability, the ongoing challenge in machine learning is to build generalized models. As a result, in order to improve future activity recognition development, we aim to explain in this paper when generalized models for feeding gesture detection fail.

The purpose of this paper is to show both how generalized feeding gesture detection models perform when trained and tested in the field, and to explore the barriers of gesture recognition in the field. We use only one wrist-worn inertial sensor on participants, and employ a two-step framework to detect and characterize eating episodes. First, we utilize an activity level recognition detector to single out the time segments when participants are sitting. We then use a motif-based time fusion classifier (MTFC) to detect feeding gestures. Finally, we test our algorithm on eight participants from an in-the-field dataset and categorize the false alarms to reinforce future machine learning model development.

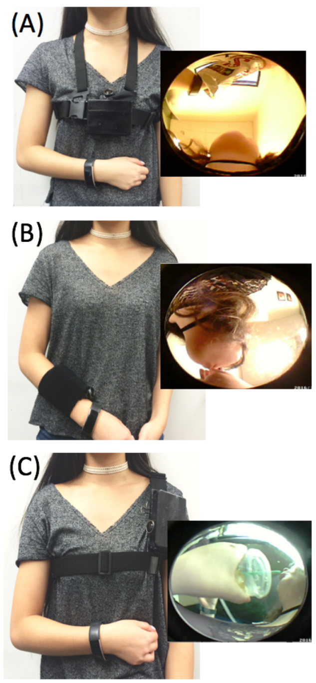

Figure 1: A participant wearing a wrist-worn sensor, a neck-worn sensor, and one of the three types of cameras: (A) chest camera, (B) wrist camera, and (C) shoulder camera

Related Works

Dong et al. show reasonable correlation between the number of bites and caloric intake in a meal in their previous work [5, 6]. Several devices with embedded sensors are being developed and deployed to detect feeding gestures, such as smart wristbands, smart necklaces, smart rings, etc [7, 14, 6, 3, 10, 4, 16, 12, 2].

Thomaz et al. employed a similar commercial wrist-worn sensor (Pebble) to detect eating episodes with a sliding window classification and clustering approach. Using accelerometer data, they showed the effect of selecting different essential parameters in their framework [14]. Dong et al. [6] showed a method for detecting an eating episodes throughout the day in free-living populations using the assumption that meals begin and end in elevated activity (since a participant often moves before and after their meal).

Literature in detecting feeding gestures has not advanced beyond 75%-80% F-measure with only wrist-worn sensor [3, 6, 14]. Prior literature has not been able to effectively analyze the reasons behind false alarms because of lack of proper ground truth data (no previous study has used a video camera in the field). The majority of ground truth in the field is generated by self-reported data. Thus our study is unique in its having video-recorded data of participants in the field, lending greater insight and discovery into why our machine learning models yield such high false alarm rates.

The contributions of this paper are three-fold. Firstly, we show results for detecting feeding episodes with a motif-based machine learning technique for eight participants in the field. Secondly, with the motif-based method, we discovered confounding gestures in free living environments that are similar to feeding gestures, providing greater insight into why our activity recognition models fail in the field. Finally, we provide categories of hand movements that result in false alarms, as well as context-based categorization of the false alarms in order to gain a better understanding of the context in which these models fail. We believe that while video cameras create privacy concerns, they can greatly increase our ability to build reliable activity recognition models that work well in the field.

Experimental Setup

To understand the barriers and facilitators to wearable adherence, we previously conducted an experiment (N=24) [1] where we asked participants to wear one of the three types of cameras (shoulder, wrist, or chest), a Microsoft Band 2 (mBand 2) wrist-worn sensor, and neck-worn sensor as shown in Figure 1. We asked them to do some structured and unstructured activities in the field and afterward we conducted interviews to understand privacy, comfort, and stigma concerns that can emerge from wearing these sensors. In this paper, we use the accelerometer and gyroscope data from the wild, along with the ground truth obtained from the camera to build feeding gesture detection models. We have processed a total of 1920 minutes of video in the study.

In this study, we found that 81.5% of feeding gestures happen with the dominant hand. Thus, with one wrist-worn sensor worn on the dominant hand, we can only detect 81.5% of feeding gestures. With wrist-worn sensors worn on both hands we could theoretically detect all the gestures, however, this would further increase participant burden (many participants are not willing to wear two wrist-worn sensors).

Data Collection

We use an mBand 2 to collect inertial sensor (including accelerometer and gyroscope) data at the frequency of 31 Hz (from options of 16 Hz, 31 Hz, and 62 Hz), and prior literature [15] found 31 Hz to be sufficient resolution in capturing feeding gesture counts. We set the frame rate of video camera at 10 Hz to ensure that the video camera can collect data for an entire day (up to 24 hrs).

One of the challenges in processing data generated from commercial wearable sensors is that the reliability of the sensor data collected fluctuates over time. To detect a feeding gesture, the reliability of the sensor data needs to be high. We estimate reliability by analyzing the sample rate on a second by second basis. We assign a reliability score between 0 and 1 for each second. If the data points collected within a second are equal to or greater than the sample rate (32 samples), the score for this second will be 1; if only half the sampling rate was collected (16 samples) the score will be 0.5. We then calculate reliability of the data for each participant by averaging the scores across the entire data set.

$$R = \frac{\sum_{t=T_{start}}^{T_{end}-1} f(1 - \frac{N_{t,t+1}}{F_s})}{T_{end} - T_{start}}$$

$$f(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

As the function in side bar shows, F_s is the frequency setting, $N(t, t+1)$ is the number of points from t to t+1. All the time variables are in seconds.

Average reliability score for the 8 participants is 96.1%, variance is 0.32%. Participants with lower than 70% reliability (2 out of 10 participants) were not included in the data analysis.

To analyze the video footage, we deploy trained annotators to label the ground truth of each participants' activity.

Methodology

Data Preprocessing

To distribute the white noise from the signals across the frequencies we apply a Gaussian kernel (sigma is set to 0.1 and the window size is set to 10). Since participants were given the ability to delete any video footage, we identify the portions of the video that were deleted and subsequently delete (after smoothing to maintain continuity of data) the corresponding inertial sensor data. Time synchronization between the inertial sensor and video camera data is a subtle but critical step in time series ground truth labeling of data. Because of the existing unpreventable time difference between the time stamp from video recording and the time stamp from the wrist-worn sensor data (which is synchronized with the smartphone clock), we recorded the smartphone's local time with the video camera to visually identify the time difference for that day. Using the recorded time in the video and the time difference between devices, we were able to associate gesture labels with labels from video.

After preprocessing, we apply a two-stage approach to eating detection that comprises motion detection (MD) and feeding detection (FD).

Two-stage Feeding Detection Approach

From the video recordings, we realized that the majority of meals were consumed when the participants were stationary. As a result, we applied motion detection first to filter out the segments of data where the participants are walking or physically active. We subsequently applied a motif-based search and classification algorithm [17] to identify feeding gestures. Our framework is depicted in Figure 2.

Motion Detection

We generated labels of the data that signal when the participants are stationary or moving. A stationary period is comprised of sitting or standing without body motion and a moving period is comprised of walking or running. For motion detection we apply a sliding window approach to segmentation of the data.

We use a simple statistical features set including means,

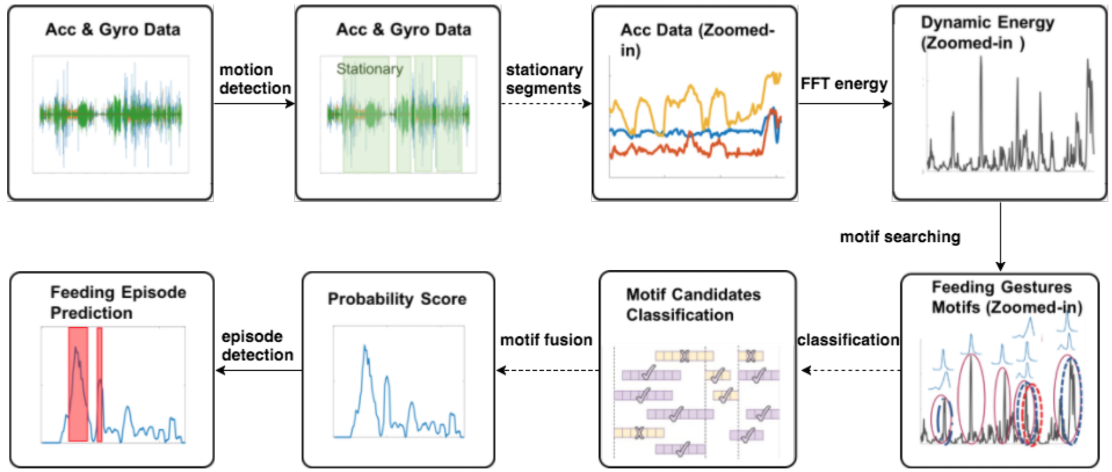

Figure 2: Feeding episode detection algorithm overall architecture

variance, kurtosis, skewness, and root mean square extracted from the 3-axis accelerometer and 3-axis gyroscope, in order to make the pre-selection step fast and feasible.

Feeding Detection
The motif-based approach we designed in a prior effort includes two main processes: motif searching to generate candidates and candidate classification. Since feeding gestures across people share similar patterns, we cluster the motifs, select the most common motifs within each cluster, and search for each one in the signal to generate candidate feeding gestures. Then we use a trained machine learning model to classify between feeding and non-feeding gestures. This approach achieves both a coarse-grained search to pick up candidates and a fine-grained classifier to recognize real feeding gestures.

Gesture Detection Step 1: Generate Dynamic Energy Signal
Selecting the type of motif to use in our search for candidate feeding gestures is essential. Since feeding gestures involve an increase in acceleration when moving the hand to the mouth, followed by a deceleration when approaching the mouth, and the reverse effect when moving the hand away from the mouth, we used the intensity or energy of acceleration to search for feeding gesture candidates.

To calculate the energy of acceleration, we first remove the effect of gravity on acceleration by removing the fundamental frequency component from our equation. $E(acc)_i$ is the dynamic energy of acceleration at time point i. $X_{x,i}$, $X_{y,i}$, and $X_{z,i}$ are the amplitudes of the frequency component of x-, y-, and z-axis acceleration. N is the Fast Fourier Transform (FFT) window size. F_s is the sampling rate.

$$E(acc)_i = \sum_{a=x,y,z} \sum_{n=2}^{N} X_{a,i}^2 (f = \frac{(n-1)F_s}{N})$$

Gesture Detection Step 2: Motif Generation
To identify the main motifs used in the search, we cluster all the motifs together using the K-Spectral Centroid Clustering (KSC) approach. The KSC method can group segments of time series data and return the generated centroid of each group. This method is based on matrix decomposition to compute the centroid of a cluster, using a distance measure for pairwise scaling and shifting.

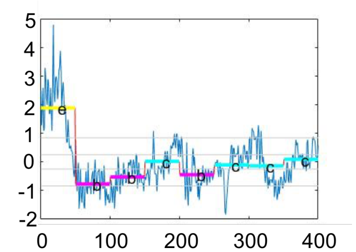

Figure 3: Example SAX representation of 400 points raw time series data (n=400, w=8, a=5)

Gesture Detection Step 3: Motif Matching and Segmentation
To further reduce the computation load for motif matching, we use Symbolic Aggregate approXimation (SAX) to reduce the dimensionality of the signals and speed up the search for motifs. SAX takes two steps to transform time series data into symbolic representation. Firstly, it normalizes the data and applies the Piecewise Aggregate Approximation (PAA) method. By equally slicing n-length time series data to w-pieces and calculating the mean value for each segment, the n-length time series data is converted to w-length real value representation. Secondly, the w-length series is transformed into symbolic representation with a dictionary of alphabet size a.

In Figure 3, the original data and SAX representations are shown as well as the intermediate representation of eight continuous data points. With the SAX method, the 400-length time series data is translated into the string 'ebbcbccc'. We take standard motifs from KSC clustering as motif templates to discover motif candidates. Then the candidates are classified by a Random Forest model.

$$pitch = arctan(\frac{G_y}{\sqrt{G_x^2 + G_z^2}})$$

$$roll = arctan(\frac{-G_x}{G_z})$$

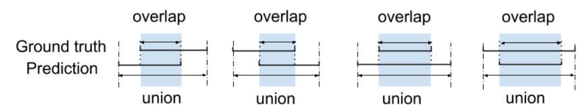

Figure 4: A diagram of measurement criterion

Gesture Detection Step 4: Feature Extraction and Classification
Since we are only using a single motif-based search approach, naturally there will be many confounding candidates identified in the search (high recall, but low precision). As a result we extract 89 features from the 6-axis signals (accel and gyro) and we calculate two other signals: pitch and roll, whose functions appear in the sidebar (shown to be useful in identifying gestures [11]) yielding 8 signals:

G_x, G_y, and G_z represent x, y, and z-axis readings from the gyro signal. We extract from the 8 signals the following features: mean, median, max, min, standard deviation, kurtosis, interquartile range, quartile 1, quartile 3, skewness, and root mean square (RMS). We also calculate the duration of time for each candidate. We then apply a trained (trained on data from other participants) Random Forest model (N=185) to classify each candidate into a feeding/non-feeding gesture classification. Since we use multiple motif searches, the identified candidates overlap in time. By summing up the number of overlapping motifs we obtain a score (representing the probability of a feeding gesture) for each time point. The higher the probability score for one time point, the greater the likelihood of it belonging to a real feeding gesture.

Gesture Detection Step 5: Defining the Event-based Measurement

Since episode prediction produces results in the form of duration, which is different from sample-level prediction, we use an event-based evaluation method to determine if a segment is true or false. When the prediction has no overlap with any ground truth, then the prediction undoubtedly fails. When there is overlap between prediction and ground truth, we categorize all the conditions into four groups: ground truth start time is ahead of prediction start time, prediction start time is ahead of ground truth start time, prediction start and end time covers ground truth, and ground truth covers the prediction. We use the equation below to decide if this prediction is true. G_1, G_2, P_1, and P_2 are start time of ground truth segment, end time of ground truth segment, start time of prediction segment, and end time of prediction segment, respectively.

$$k = \frac{min(G_2, P_2) - max(G_1, P_1)}{max(G_2, P_2) - min(G_1, P_1)}$$

To determine if a candidate's feeding gesture was correct or not, we used event-based evaluation, where we defined an overlap ratio between ground truth and the candidate's feeding gesture to decide whether a segment was correctly identified or not. We defined overlap, k, as the ratio of the overlap of ground truth and prediction segments to the combination of both ground truth and prediction segments.

Results and Analysis

Motion Detection

To ensure we captured all feeding episodes, we opted for a high recall classifier in detecting motion. We selected a sliding window size of two minutes and a stride of one minute. We also added a one minute padding to the start and end of each stationary episode identified, to avoid missing any stationary episodes with feeding. We used Leave One Participant Out (LOPO) approach to train the motion model and predict the stationary episode. The LOPO motion detection achieves an average 98.25% recall and an average 60.1% precision.

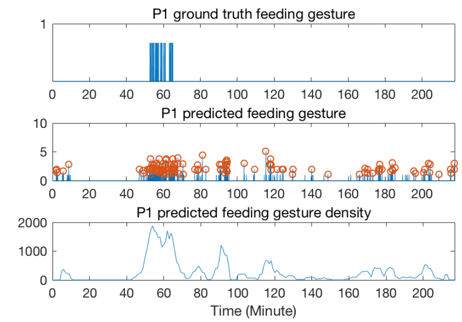

Figure 5: Feeding episode prediction result for Subject 1: (a) shows the ground truth feeding gesture duration, (b) shows the score for each time point from prediction, red circle represents peak, (c) shows the density of the predicted feeding gestures.

Gesture Detection

To generate feeding gesture motifs, we used the KSC clustering and SAX method to generate energy-based motif templates from the training set using ground truth. Empirically we obtained optimal performance using the SAX method with alphabet size a = 5. Given a motif length n equal to the length of time series based on ground truth, the number of symbols in the low dimensional approximation of the sub-sequence is set to $w = n/2$.

Searching through the signal using the centroid motifs generates several candidate feeding gestures (high recall as shown in Table 2), based on an overlap criterion (shown in Figure 4) of 50%. The smaller the criterion, the less required overlap between ground truth and the predicted signal, and the higher the recall.

Gesture	Count
texting on phone	462
walking	237
drinking	94
turning a page	58
writing	51
gesturing while talking	35
fidgeting with hand	33
holding onto train pole	28
adjusting camera	29
lifting food without eating	18
scratching eye	22
scratching/covering/touching mouth	21
ordering food	18
touching face	17
wiping mouth	15
wiping face	14
scratching/touching chin	11
scratching/touching nose	10
scratching/touching head	14
lifting cup without drinking	14
opening a door	14
stirring noodles	13
adjusting glasses	13
biting nails	10
packing	9
capping pen	8
raising hand	8
picking up and putting down phone	8
unwrapping food	7
wiping off bookshelf	7
picking up stuffs	7
wiping hands	7
touching hair	7
fidgeting with food	5

Table 1: Categorization result for false alarm gestures

After we identify the candidate feeding gestures, we summed the density of the feeding gestures using a 5-minute window, as shown in Figure 5. We show the LOPO feeding episode prediction result for the eight participants in Table 2.

False Alarm Analysis
In this work we further explore the reasoning behind the high false alarm rate of feeding gesture detection. We analyzed the video and checked the false alarm moments to categorize the confounding gestures that caused false alarms. We show a detailed distribution of all the gestures from the in-the-field real life test in Table 1.

As shown, texting on the phone causes a considerable amount of false alarms. When we checked the video we found that several false alarms occurred when participants were lying on the couch and raising their hands to play with their phones. Because we trained a LOPO generalized machine learning model on seven subjects, and tested on one, high false alarms occur with subjects who have unique feeding and confounding feeding gestures. This could potentially be solved by having a larger sample size with more feeding gestures.

We categorized confounding gestures shown in Table 1 into four categories: hand up and down, extending the hand, complex hand movement, and moving vibrations. We further categorized the context of the false alarms as shown in Table 4, with texting on the phone producing the largest percentage of false alarms.

From our results we find that the high false alarm rates can be attributed in part to incorrect wearing of the device. When the watch is not worn with the face facing upward, horizontal hand gestures are recognized as vertical gestures. Adhering to proper wearing of the technology is a challenge, and greatly impacts the ability to capture true feeding gestures.

Subject	recall(p)	recall(n)	precision(p)	precision(n)	f1(mean)
P1	1	0.93	0.48	1	0.8
P2	1	0.88	0.48	1	0.79
P3	0.5	0.77	0.08	0.97	0.5
P4	1	0.46	0.12	1	0.43
P5	0.67	0.68	0.2	0.94	0.55
P6	1	0.52	0.21	1	0.51
P7	1	0.51	0.15	1	0.47
P8	1	0.88	0.5	1	0.8
Ave	0.90	0.70	0.28	0.99	0.61

Table 2: Feeding gesture detection metrics by minute (p for positive and n for negative)

Category	Example	Percent
Hand up and down	touching facial area, covering mouth, wiping mouth, scratching nose, scratching face, picking up sensor from table, moving glass up and down, flipping page, biting nails, rub mouth,picking something up	23%
Extending hand	put food on table, picking up sensor from table, grabbing food/drink	6%
Complex hand movement	texting on phone, wiping hand, playing with bottle, unwrap food, fidgetting while sitting ,wiping hands	53%
Moving vibration	walking or by vehicle	18%

Table 3: Categorization result for false alarm gestures

Conclusion and Future work
Recent advancements in embedded wearable devices have made it possible to passively identify eating by detecting feeding gestures with the use of on-body inertial sensors.

Category	Example	Percent
phone related gesture	texting on phone	38%
transportation	walking,	21%
facial area gesture	scratching/touching chin/eye/nose/mouth, wiping mouth	13%
paper related gesture	flipping paper, writing	9%
communication gesture	gesturing while talking, raising hand, ordering on menu	5%
food related gesture	lifting food without eating, stirring the noodle, unwrap food	4%
fidgeting	capping pen, fidgeting with hand	4%
others	packing, opening the door, pick up something, wiping off bookshelf	6%

Table 4: Context categorization result for false alarm gestures

In this work, we applied a motif-based feeding gesture detection system and tested it in an in-field study yielding high recall, but a high false alarm rate, resulting in an average F-measure of 61%. We further categorized the false alarms into different categories, with complex hand movements yielding the greatest percentage of false alarms, and phone-related gestures (texting on the phone) yielding the largest percentage (38%) of confounding gestures. For future work, we aim to modify our motif-based approach to handle each of the false alarm categories identified in this paper.

REFERENCES

1. Rawan Alharbi, Angela Pfammatter, Bonnie Spring, and Nabil Alshurafa. 2017. WillSense: Adherence Barriers for Passive Sensing Systems That Track Eating Behavior. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2329–2336. DOI: http://dx.doi.org/10.1145/3027063.3053271

2. Nabil Alshurafa, Haik Kalantarian, Mohammad Pourhomayoun, Jason J. Liu, Shruti Sarin, Behnam Shahbazi, and Sarrafzadeh Majid. 2015. Recognition of Nutrition Intake Using Time-Frequency Decomposition in a Wearable Necklace Using a Piezoelectric Sensor. *IEEE Sensors Journal* 15, 7 (July 2015), 3909–3916. DOI: http://dx.doi.org/10.1109/JSEN.2015.2402652

3. Oliver Amft, Holger Junker, and Gerhard Troster. 2005. Detection of Eating and Drinking Arm Gestures Using Inertial Body-worn Sensors. In *Proceedings of the Ninth IEEE International Symposium on Wearable Computers (ISWC '05)*. IEEE Computer Society, Washington, DC, USA, 160–163. DOI: http://dx.doi.org/10.1109/ISWC.2005.17

4. Oliver Amft and Gerhard Tröster. 2009. On-body Sensing Solutions for Automatic Dietary Monitoring. *IEEE pervasive computing* 8, 2 (2009), 62–70.

5. Yujie Dong, Adam Hoover, Jenna Scisco, and Eric Muth. 2012. A New Method for Measuring Meal Intake in Humans via Automated Wrist Motion Tracking. *Appl Psychophysiol Biofeedback* 37, 3 (Sep 2012), 205–215.

6. Yujie Dong, Jenna Scisco, Mike Wilson, Eric Muth, and Adam Hoover. 2014. Detecting Periods of Eating During Free-Living by Tracking Wrist Motion. *IEEE Journal of Biomedical and Health Informatics* 18, 4 (July 2014), 1253–1260. DOI: http://dx.doi.org/10.1109/JBHI.2013.2282471

7. Holger Junker, Oliver Amft, Paul Lukowicz, and Gerhard Tröster. 2008. Gesture Spotting with Body-worn Inertial Sensors to Detect User Activities. *Pattern Recognition* 41, 6 (2008), 2010 – 2024. DOI: http://dx.doi.org/10.1016/j.patcog.2007.11.016

8. Christos Maramis, Vasilis Kilintzis, and Nicos Maglaveras. 2016. Real-time Bite Detection from Smartwatch Orientation Sensor Data. In *Proceedings of the 9th Hellenic Conference on Artificial Intelligence (SETN '16)*. ACM, New York, NY, USA, 30:1–30:4. DOI:http://dx.doi.org/10.1145/2903220.2903239

9. Katherine M. Flegal; Margaret D. Carroll; Brian K. Kit; Cynthia L. Ogden. 2012. Prevalence of Obesity and Trends in the Distribution of Body Mass Index Among US Adults, 1999-2010. *JAMA* 307, 5 (2012), 491–497. DOI:http://dx.doi.org/10.1001/jama.2012.39

10. Tauhidur Rahman, Mary Czerwinski, Ran Gilad-Bachrach, and Paul Johns. 2016. Predicting "About-to-Eat" Moments for Just-in-Time Eating Intervention. In *Proceedings of the 6th International Conference on Digital Health Conference (DH '16)*. ACM, New York, NY, USA, 141–150. DOI: http://dx.doi.org/10.1145/2896338.2896359

11. Nazir Saleheen, Amin Ahsan Ali, Syed Monowar Hossain, and Hillol Sarker. 2015. puffMarker: A Multi-sensor Approach for Pinpointing the Timing of First Lapse in Smoking Cessation. In *Proceedings of the 2015 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 999–1010.

12. Sougata Sen, Vigneshwaran Subbaraju, Archan Misra, Rajesh Krishna Balan, and Youngki Lee. 2015. The Case for Smartwatch-based Diet Monitoring. In *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. 585–590. DOI: http://dx.doi.org/10.1109/PERCOMW.2015.7134103

13. Liz Szabo. 2016. Federal Government Urges Americans to Limit Sugar to 10% of Daily Calories. (2016). https://www.usatoday.com/story/news/2016/01/07/federal-dietary-guidelines/77151060/

14. Edison Thomaz, Irfan Essa, and Gregory D. Abowd. 2015. A Practical Approach for Recognizing Eating Moments with Wrist-mounted Inertial Sensing. In *Proceedings of the 2015 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 1029–1040.

15. Shibo Zhang, Rawan Alharbi, Mohammad Pourhomayoun, William Stogin, and Nabil Alshurafa. 2016a. Food Watch: Detecting and Characterizing Eating Episodes through Feeding Gestures. In *International Conference on Body Area Networks, BodyNets 2016*. Italy.

16. Shibo Zhang, Rawan Alharbi, William Stogin, Kevin Moran, Angela F. Pfammatter, Bonnie Spring, and Nabil Alshurafa. 2016b. Machine Learning Algorithms Applied to Detect Feeding Gestures. *The Obesity Society* (2016).

17. Shibo Zhang, William Stogin, and Nabil Alshurafa. to be published. I Sense Overeating: Motif-based Machine Learning Framework to Detect Overeating Using Wrist-worn Sensing. *Information Fusion* (to be published).