Are Videos Reliable? Deciphering Problems with Video-based Ground Truth in Free-living Studies

ABSTRACT

Wearable mHealth (mobile health) devices can monitor various activities and behaviors performed by individuals in both laboratory and free-living settings. To evaluate the performance of mHealth devices, researchers often compare the device's output against collected ground-truth (e.g., data from diet interviews, video recordings, etc.). However, different ground truth collecting approaches can introduce unique biases that impact the data collection and analysis process. In this paper, using diet monitoring activity as an exemplar, we describe our experience in collecting over 630 hours of video-based ground truth from 15 participants in free-living conditions, and the challenges we faced in utilizing this ground truth for data analysis. We observed that annotators could not identify 23.15% of the food items that were captured in the video-recordings. Using specific examples from our diet monitoring experience, we provide recommendations for how future researchers can improve approaches to capturing video-based ground truth about activity monitoring.

Author Keywords

eating activity monitoring, video-observations, self-reports

CCS Concepts

•**Human-centered computing** → **Human computer interaction (HCI);** *Haptic devices;* User studies; Please use the 2012 Classifiers and see this link to embed them in the text: https://dl.acm.org/ccs/ccs_flat.cfm

INTRODUCTION

Obesity is a major public health concern in the United States and is associated with many chronic diseases, such as cardiovascular disease, diabetes, and certain cancers [14]. To create interventions that encourage healthy behaviors for promoting weight loss and reducing obesity risks, it is important for clinicians and researchers to understand an individual's eating habits (e.g., eating patterns, food choices) and physical activities.

With the growing popularity of smartphones and wearable mobile devices, monitoring various aspects of everyday activities and behaviors, including mHealth, is becoming prevalent. These devices allow automatic and unobtrusive monitoring of various health characteristics in free-living settings. Some common mHealth use-cases of these mobile and wearable devices include using a fitness-tracker to monitor the number of steps that an individual takes [6], using a smartphone to monitor sleep patterns [4], or using a chest band to monitor the individual's heart-rate and stress levels [9]. However, numerous mobile and wearable based mHealth prediction systems are still in the research phase and researchers aim to validate these systems in free-living studies against established ground truth. In this paper, we specifically focus on a mHealth system that we have designed and developed. The objective of the system is to monitor eating and dietary monitoring in free-living setting. In this paper, we do not focus on the sensor details, but rather focus on the ground truth that we collected for monitoring an individual's eating behavior. Ground truth can be determined using various methods; however, the wearable camera is emerging as a popular tool to capture ground truth activity data. Researchers and clinicians have used images or videos captured by these mobile or wearable cameras to predict activities or behaviors (i.e., as a sensor) by employing innovative image-processing techniques or by confirming activities and behaviors reported by participants or wearable sensors. Wearable cameras are well-positioned to advance health behavior monitoring, particularly with diet assessments. Figure 1 presents images captured by cameras in various research works for either predicting activities or for capturing the ground truth.

Historically, data from self-reported dietary intake data has been considered the ground truth for information about eating behaviors. One common method to capture self-reported diet data is the 24-hour diet recall, a type of diet interview that relies on the individual's memory to report any food and beverage consumption within the last 24 hours. However, this approach is subject to reporting bias and individuals often either over- or under- report their calorie intake and food consumption [10]. Wearable cameras can provide objective data that can address the reporting bias from 24-hr diet recalls and confirm actual intake, although they can have their own set of limitations as well.

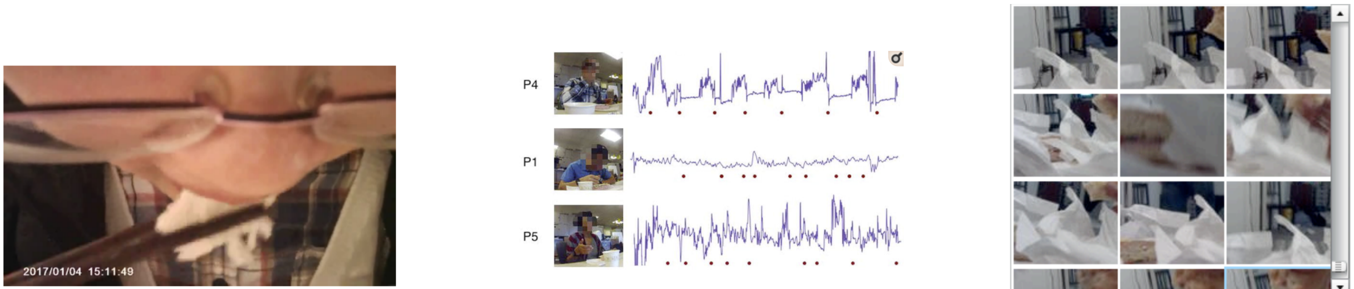
Previous literature have used video cameras to record an individual's eating activity in laboratory setting [18]. However, behaviors demonstrated by individuals in a laboratory setting are typically structured and may not be reflective of usual health behaviors demonstrated by the same individual in free-living conditions. Thus, it is unclear whether data collected from wearable cameras are consistent with those collected from 24-hr diet recalls in free-living settings.

In this paper we focus on using the video-camera as a ground truth monitoring device. Herein, we aim to empirically answer the following question "Does the video-observations captured by a camera help provide detailed ground truth?". To answer

CHI'20, April 25–30, 2020, Honolulu, HI, USA

ISBN 978-1-4503-6708-0/20/04. . . $15.00

DOI: https://doi.org/10.1145/3313831.XXXXXXX



(a) Ground truth as captured from a hat by Bi et al. [2]

(b) Ground truth as captured from a camera placed in front of participant by Thomaz et al. [17]

(c) Images captured from a POV camera by Reddy et al. [15]

Figure 1: Images (or frames from videos) captured by researchers in eating activity monitoring studies. These studies either used the images to predict the activities or to capture the ground-truth

this question, in this paper we analysed 630 hours of video data that we collected from 15 participants, across 144 days in a free-living study to monitor the differences in eating behaviors of individuals with and without obesity. We identify and systematically categorize various factors that we observed in the video recording that affected annotating the food items that participants consumed.

The main contributions of our work includes:

1. Empirically demonstrating that video-observations might not always provide accurate ground truth. In a 144 person-days study with 15 participants who consumed 514 food items, we observed that annotators could not identify 23.15% of the food items that were captured in the video-observations. As compared to 24-hour diet recall, we identified that 60.6% of food items were not captured by video-observations.

2. Investigating and systematically categorizing various factors that hampered the item identification process. Overall, we group these factors into three categories. For each category, we provide possible recommendations to improve the collection of video-observation diet data.

BACKGROUND AND RELATED WORK

Self-reports, sensor-measurements and video-observations are three common approaches for monitoring diet and physical activities in free-living conditions. While researchers can independently analyze image- and video-observations to capture usual diet and physical activity behaviors in the context in which they are occurring (e.g., setting, time) [3, 5, 15], they can also use the image- and video-observations as the ground truth when evaluating other monitoring and prediction techniques [2, 11, 17].

Several researchers have utilized self-reported data as ground truth for monitoring or analysing health behaviors [7, 13]. However, self-reported diet data are affected by various factors – e.g., social norms can influence individuals to over- or under- report calories, [10], or participants might forget to periodically self-report, thus introducing the problem of missing-data [8]. Video camera recordings can address the limitations of self-reported data but faces other challenges, including image clarity, image obfuscation, and privacy concerns of participants [1]. In the past, researchers have compared the three methods of assessing activities or behaviors with one another, i.e., self-reports, video-observations, and sensor-measurement, to determine the best approach to capturing true activity data [16], and they determined that video-observations were the most precise method in their environment. However, the researchers did not pin-point scenarios where the validity of video-observations for activity monitoring can be impacted by varying factors. In this paper, we do not consider sensor-measurement to be ground truth (considering that various mobile health (mHealth) systems are still in research phase), but rather aim to validate the measurements-reports using the video-observations and the self-reports. Using a dataset that we collected, we identify and categorize challenges in using video-observations as a dietary assessment.

DATA COLLECTION AND ANALYSIS

We next describe the dataset and subsequent analyses that we performed.

The eating activity monitoring dataset

This study was part of a larger IRB-approved project to monitor differences in eating habits of participants with obesity and participants without obesity in a free-living setting. Of the overall sample, 15 participants volunteered to collect video data (9 obese, 6 non-obese). Participants were aged between 19 and 62 years ($\bar{x} = 33.4$ years, $\sigma = 13.14$ years). The recruited participants were provided with multiple devices that they had to wear for a period of 2 weeks. Specifically, these devices included (a) a wearable neck-worn device to automatically monitor the eating activity from multi-sensor measurements, (b) a wearable camera with a fish eye lens that was strapped on the non-dominant shoulder, (c) a smartphone that aggregated the data from the necklace and the camera. The neck-work sensor continuously monitored the eating activity, while the video camera captured the video-observations through the day. In this paper, we do not focus on the neck-worn device's data, but rather focus on the video camera's output to determine its performance. In addition to the video recording, a registered dietitian performed daily 24-hr diet recalls with each participant over telephone using a multiple

#	Collection days	Total items in video	Indescipher-able items
P01	10	40	9
P02	6	27	5
P03	14	56	17
P04	13	29	9
P05	8	30	1
P06	7	16	5
P07	2	7	1
P08	11	44	9
P09	10	44	23
P10	14	43	16
P11	14	58	2
P12	11	44	7
P13	5	23	3
P14	12	30	8
P15	7	23	4
Total	144	514	119

Table 1: Participant wise distribution of detected and indecipherable items in the videos. P1-P9 were participants with obesity, while P10-P15 were participants without obesity.

pass method and Nutrition Data System for Research (NDSR) software, a common approach used to collect self-reported diet data in nutrition research. Participants were asked to recount all the food and beverage items that they consumed in the last 24-hours.

Analysis

Table 1 presents a distribution of the number of days that each participant participated in our study, along with the distribution of food items that they consumed during the study. We collected a total of 144 person-days of video data from the study sample. We employed video annotators to review the video data. Overall, the annotators identified 395 food items during the 144 person-days of video-observations, while there were at least 119 (23.15%) food items that were unclear in the videos and deemed *indecipherable*. We next analysed video snippets to identify potential causes of indecipherable food items.

Six main causes emerged when reviewing the videos of indecipherable food items: (i) *obstructed* – another object blocked the camera from capturing an image of the food item , (ii) *out of view* – the food item was not in the camera's field of view, (iii) *lens flare* – lens flare obstructed view of the food item, (iv) *lighting issues* – scene was either too dark or too bright, (v) *image quality issues* – food item was pixelated, and (vi) *unknown food item* – annotators could not recognize the food item that was consumed. Figure 2 presents the distribution of each of the 6 categories in our videos. As evidenced in the figure, the primary factor that caused food items to be indecipherable was because of foods not being in the lens' field of view.

To answer the question whether video-cameras provide detailed ground truth, we compared food data identified in the video-observations against the information in the 24-hr dietary recalls. We observed that there was a 39.4% overlap

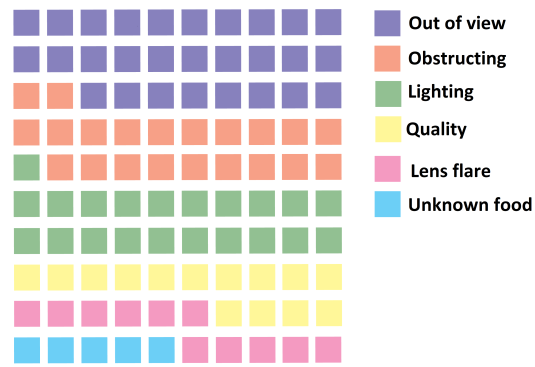


Figure 2: High level categorization of the indecipherables

between the items reported in the self-report and items identified in the videos. We also observed there were 791 items that were present in the self-report, but were not detected in the video-observations.

CHALLENGES IN IMAGE BASED GROUND TRUTH

We identified six causes for indecipherable food items – i.e., instances where the annotators could not identify the food items in the video. These six causes were then grouped based on three types of factors – instances affected by human factors, instances affected by environmental factors, and instances affected by the video-capturing device (device factors). Table 2 presents the three overarching factors that caused indecipherable food items.

(i) address food placement with the participant at study baseline, (ii) discuss the importance of capturing usual intake with the participant, and emphasize that there is no judgement, (iii) use subjective and objective diet data to provide a comprehensive picture, (iv) train annotators on how to search for unfamiliar food items

C1. Instances affected by human factors

In an in-lab control study, a member of the research team is usually available to ensure that the participant's actions does not affect the video-recording. However, we observed that there were several human factors that affected the video recording in a free-living setting. Additionally, once we extracted the video recordings, we observed that sometimes issues crept in due to inconsistency caused by the video annotators.

a. Participant obstructs the camera:

One major challenge in identifying items from a video-observation was when the items were not captured by the video camera –i.e., item was obstructed by the participant or by objects. We observed that obstruction or blocking occurred when (i) an obstructing object (e.g., a bowl or a tumbler) was placed between the food item and camera. The participant might have placed this obstructing object while performing a natural action – e.g., after drinking from a cup; or (ii) the participant was performing a natural eating gesture and unintentionally blocked the camera – e.g., lifting a bowl of food

#	Factor	Sub-category	Occurrences	Recommendations
C1	Human Factors	Obstruction Deviating from normal Turning OFF camera Annotation issues	Frequently Often N/A Rarely	(i) Address food placement with participant at study baseline (ii) Discuss the importance of capturing usual intake with participants, and emphasize that there is no judgement (iii) Use subjective and objective diet data to provide a comprehensive picture (iv) Train annotators to search for unfamiliar food items
C2	Environmental Factors	Lighting Factor Lens flares Physical location	Frequently Often Often	(i) Trigger infrared (IR) sensor as necessary (ii) Notifying participant to adjust position of camera (iii) Review self-reports to provide additional information
C3	Device Factors	Capturing resolution Device position Device components Motion blur	Often Frequently Often Rarely	(i) Adapt the capturing resolution based on scenario (ii) Allow subtle camera rotation to capture items of interest (iii) Reduce automatic configurations and adapt manually Securely strapping the camera

Table 2: Challenges and possible recommendations for capturing activity and behavior monitoring videos in free-living conditions.

and bringing it very close to the mouth; or (iii) the item was consumed in an opaque container – e.g., a beverage in an opaque bottle. In future, when such a situation occurs, we recommend either nudging the participant to inform them that a certain object is blocking the field of view, or use surrounding video context (timing, environment, eating utensils) and compare with self-reported dietary data to confirm whether the food was actually consumed by the participant.

b. Participant deviating from the expected:
All video-capturing devices have a specific field of vision. During user studies, researchers position these devices so that it can capture images (or video) of the item of interest. For example, Figure 1a presents the images captured from a camera mounted on the under-brim of a hat by Bi et al [2], while Figure 1c presents images captured by Reddy et al. using a camera that is suspended from the neck using a lanyard [15]. Researchers generally instruct participants to consume food while ensuring that the process occurs within the field of view of the camera. However, unlike controlled in-lab studies, researchers do not have control over the participant's actions, be it in terms of food choices, sitting positions or even gestures that they perform; participants should not be controlled in a free-living study. In our study, we observed that our camera could not capture images when participants turned their head and consumed food with the non-dominant hand. Herein, a potential solution will be to design and develop innovative mechanical design that allows the camera to rotate and adapt based on the location of the food item.

c. Participant turning off the device:
There can be many reasons why participants decide not to record their activities, including privacy issues or feeling uncomfortable about their actions. We observed several data gaps in our video recordings and we do not have any video-based information to determine whether we missed monitoring an eating activity during that period. However, per participants' 24-hr diet recalls, it appears that we missed several meals during the 2-week period. In fact, we identified that we did not capture over 60% of the consumed food items that were reported in the self-report. Although we are currently relying on participants to accurately record the video based ground truth, yet we observe several instances when participants pause video capturing. In future we can either adopt innovative obfuscation approaches, or trigger occasional Just-in-Time Adaptive Interventions(JITAIs) when participants switch off the camera.

d. Annotator uncertain about item:
At the end of the annotation process, we informally interviewed the annotators to determine challenges that they faced during the annotation process. The annotators raised two interesting insights: (i) they were unfamiliar with types of food items eaten across different cultures, which impacted how well they could identify food items and (ii) sometimes an annotator might observe that participants were consuming food. However, they could not see what the participant was consuming. If the annotator had a clear view during the food-preparation process, then the annotator could determine the likely food item that the participant was consuming. In future, we will better train the annotators and consider allowing them access to the food diary if necessary.

C2. Instances affected by environmental factors
In addition to human factors, numerous environmental factors can affect the video-data collection in a free-living study . In a field-study, we anticipate participants will perform the tasks-of-interest (e.g., eating) as we expect them to. Our analyses revealed that video annotators indicated that in almost 50% of cases, environmental factors made item identification challenging. More specifically, we identified the following environment factors that affected our dataset.

a. Eating in dimly-lit environments:
From the video-observations, we noted that several participants consumed food in a dimly-lit environment environment – e.g., consuming food while watching television a program. Although our camera was equipped with an IR sensor, which can capture images in darkness, the camera also captured glares from the television, which were high in intensity and thus prevented the IR sensor from switching over to night-mode. We recommend that the IR sensor mode should be selected so that the function is triggered in dimly lit environments.

(a) A frame from the video captured while the participant consumed salad.

(b) A frame from the video captured while participant consumed smoothie.

(c) (C2a) A frame extracted from the video when the participant consumed a meal while sitting in a dimly lit location. It is difficult to detect the exact food consumed during this meal.

(d) (C1a) A frame extracted from the video showing that the participant is consuming food. However, it is difficult to determine the consumed food item.

Figure 3: Frames extracted from the video captured by the ground-truth capturing camera. It is possible to identify the food item consumed in frames (a) and (b), while it is difficult to identify the food items consumed in frames (c) and (d)

b. Lens flares:

In addition to glares from the television affecting the IR sensor, we observed several video segments affected by lens flares. Light rays from a bright source can hit the lens, causing various lens elements to reflect and refract the light and in-turn, create unwanted ghost-lighting or lens flares. We observed that since we used a fish-eye lens to capture the videos, the lens flares appeared in the form of a semi-transparent circular object that was centered at the center of the scene, which impacted data collection. Some common sources of lens flares that we observed were caused when it was bright outside and the recording camera faced a window, or when the participant was in a restaurant and emissions from a focus light directly hit the camera. One possible precaution to take is to analyse the video recording output in run time to determine if it is being washed out. In such a situation, an immediate notification can be sent to the participant, informing about the disturbing light source and asking the participant to make necessary adjustments.

c. Physical Location:

Annotators indicated several eating instances that occurred at a particular location or type of location that were challenging to annotate. For example, the annotators indicated that it was difficult to decipher food consumed in a dimly lit restaurant or bar. Annotators also indicated that when they could not determine the quantity of a food item that the participant consumed, they were likely to indicate that the participant consumed one of that item. However, we found that in all instances, the number of food items consumed by participants reported in the 24-hr diet recalls was more than the average number of food items indicated by annotators from video data. At this stage, this highlights the importance of utilizing self-reporting mechanisms, in addition to video-recording observations.

C3. Instances affected by factors related to the video-capturing device

In addition to human and environmental factors, certain device-related factors also affects the video capturing. Some of these factors overlap with factors indicated in the previous sections.

a. Capturing resolution:

One major challenge with video-recording a participant's entire waking day is to ensure that the video-recorder does not run out of power. Researchers, including our research group, have applied various techniques to reduce the power consumption. One such technique is to capture images or videos at a lower resolution [12]. However, annotators found it challenging to annotate certain images or videos that were captured at lower resolution (e.g., in a dimly or extremely brightly lit location). We recommend considering an an adaptive image capturing pipeline where the capturing resolution can be adjusted depending on scenario and moments of interest.

b. Position from where device is capturing:

Researchers vary the position of the camera to adjust the camera's field of view based on the activity of interest. Since our group is interested in the eating activity, we conducted several feasibility studies to identify the best position of a wearable camera for capturing eating activity related observations. We identified that a camera with a fisheye lens that is placed on the shoulder can capture both (a) the participant's mouth and (b) the food item in front of the participant. Sample images from our device is presented in Figure 3. Although we identified that this position was optimal in a controlled setting, as indicated in *C1*, unpredictable human behavior can affect the recording. In such scenarios, either considering a mechanical design to rotate the camera or providing adaptive interventions to participants might be essential.

c. Device component limitations:

We observed that the camera usually requires time to adapt to the lighting conditions. This can be attributed to the characteristics of certain device components. Although higher quality lens might help lower this adaptation time, however, such lenses can be expensive. Future research should consider maintaining a stable camera aperture for environments where the light intensity is continuously varying. This aperture value can be determined using the average of intensity variation during a period of time.

d. Motion blur:

We found during the feasibility studies that a free-hanging camera was prone to motion blur. Thus, we decided to incorporate a design where the camera was strapped firmly onto the participants. This strategy helped reduce the number of captured instances affected by motion blur. However, we did observe motion-blur during certain instances where the participants moved, but since the movement was for a small duration, the annotators could still identify the food item during these instances. Researchers using freely suspended cameras in free-living studies might consider a design where the camera is strapped onto the participant across the shoulder.

DISCUSSION

Although we grouped the challenges into three factors, there is some overlap between them. We discuss some research questions that arose because of these challenges and identify potential solutions for these challenges.

Interventions (C1a-c, C2b-c, C3b): For our indecipherable and problematic scenarios, we observed that if we could trigger Just-in-Time Adaptive Interventions (JITAIs) at those moments, we could obtain better video-observations. However, to trigger JITAIs, it is essential to detect the problematic or indecipherable contexts. Thus, it will be interesting to evaluate possible techniques that can be utilized to identify these problematic scenarios accurately and in near-real time. Additionally, since interventions will bring the user in the loop, it is also important to ensure that the interventions does not affect usability.

Improving the camera's mechanical design (C1a-b, C2b, C3b): We had utilized a wearable camera in our study that was strapped onto the participants. However, in several scenarios, we identified that if we could subtly rotate the camera's lens, we could have avoided several indecipherable moments – e.g., a slight rotation could remove the lens from being directly in line of a light source. Although several possible innovation directions in mechanical design are possible, it will be interesting to determine these problematic scenarios in real-time and identify techniques that can be used to adapt to them in a case-by-case manner.

Training (C1a-d): While evaluating the human factors that caused the indecipherable food items, we identified several cases that could have been avoided if we instructed participants about how to use these cameras and annotators on how to interpret food items. Since we intend to collect data in free-living setting, we did not train participants about things to do or not to do outside of turning on the camera. However, we posit that if we had provided certain instructions to participants related to camera position and lighting, their usual behavior should not be hampered and we would have fewer indecipherable food items. Thus, a future research question to address might be: "*How should we train the participants to capture valid video data so that it does not affect their usual behavior?*'. In addition to participants, it might also be interesting to consider the appropriate amount of training that should be provided to the annotators.

Adaptive pipeline (C2a-b, C3a, C3c): Most traditional video-recording techniques used in free-living settings have a simple static processing pipeline. However, various various in-field scenarios necessitate using an adaptive pipeline - -e.g., capturing images at a higher resolution when *item of interest* is detected. How, then should we design the the sensing and processing pipeline to maximize the system's performance?

Augmenting video-observations with Self-reports (C1a-c, C2c, C3b, C3d): To maintain design simplicity or reducing user-burden, researchers adopt single ground truth capturing approach, i.e., either using fully automated video-recording or completely manual self-reporting. We believe that each technique by itself is not sufficient for collecting ground truth due to unique biases, especially in free living setting. In this paper, we discussed about several scenarios where annotators encountered indecipherable items in the video-observations. Relying on video-observations alone thus reduces the reliability of the collected ground truth due indecipherable food items yet relying on self-report can introduce reporting biases. An interesting research direction to investigate the most optimal way to integrate diet data from both video-capturing and self-report approaches.

CONCLUSION

In this paper, we discuss the challenges of video recording in free-living conditions. Using a dataset that we collected with 15 participants and 630 hours of data, we observed that annotators could not identify food items in 23.15% of the eating instances. In this paper, we systematically identify these indecipherable instances, the *indecipherables*, and provide recommendations that future researchers can consider when using cameras to monitor dietary intake. Although researchers rely either on video-observations or self-reports for ground truth, we believe that an approach where researchers use both video-observations and self-report will increase the validity of the data when monitoring the eating behavior or any other activities that occur in free-living conditions. Clinicians and researchers can use the data to monitor various activities, including eating, and design effective interventions, e.g., weight loss interventions to reduce obesity risk.

REFERENCES

[1] Rawan Alharbi, Tammy Stump, Nilofar Vafaie, Angela Pfammatter, Bonnie Spring, and Nabil Alshurafa. 2018. I Can't Be Myself: Effects of Wearable Cameras on the Capture of Authentic Behavior in the Wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (9 2018), 1–40. DOI: http://dx.doi.org/10.1145/3264900

[2] Shengjie Bi, Kelly Caine, Ryan Halter, Jacob Sorber, David Kotz, Tao Wang, Nicole Tobias, Josephine Nordrum, Shang Wang, George Halvorsen, Sougata Sen, Ronald Peterson, and Kofi Odame. 2018. Auracle: Detecting Eating Episodes with an Ear-mounted Sensor. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (9 2018), 1–27. DOI:http://dx.doi.org/10.1145/3264902

[3] Robert Bodor, Robert Bodor, Bennett Jackson, Nikolaos Papanikolopoulos, and Human Tracking. 2003. Vision-based human tracking and activity recognition. *Proceedings of the Mediterranean conference on control and automation* 1 (2003), 18–20. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.116.9664

[4] Zhenyu Chen, Mu Lin, Fanglin Chen, Nicholas D. Lane, Giuseppe Cardone, Rui Wang, Tianxing Li, Yiqiang

Chen, Tanzeem Choudhury, and Andrew T. Campbell. 2013. Unobtrusive sleep monitoring using smartphones. In *Proceedings of the 2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, PervasiveHealth 2013*. 145–152. DOI: http://dx.doi.org/10.4108/icst.pervasivehealth.2013.252148

[5] Felicia Cordeiro, Daniel A. Epstein, Edison Thomaz, Elizabeth Bales, Arvind K. Jagannathan, Gregory D. Abowd, and James Fogarty. 2015. Barriers and negative nudges: Exploring challenges in food journaling. In *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 2015-April. Association for Computing Machinery, 1159–1162. DOI: http://dx.doi.org/10.1145/2702123.2702155

[6] Keith M. Diaz, David J. Krupka, Melinda J. Chang, James Peacock, Yao Ma, Jeff Goldsmith, Joseph E. Schwartz, and Karina W. Davidson. 2015. Fitbit®: An accurate and reliable device for wireless physical activity tracking. *International Journal of Cardiology* 185 (4 2015), 138–140. DOI: http://dx.doi.org/10.1016/j.ijcard.2015.03.038

[7] Simon Flutura, Andreas Seiderer, Ilhan Aslan, Chi Tai Dang, Raphael Schwarz, Dominik Schiller, and Elisabeth Andre. 2018. DrinkWatch: A mobile wellbeing application based on interactive and cooperative machine learning. In *ACM International Conference Proceeding Series*, Vol. 2018-April. Association for Computing Machinery, 65–74. DOI: http://dx.doi.org/10.1145/3194658.3194666

[8] Susan M. Fox-Wasylyshyn and Maher M. El-Masri. 2005. Handling missing data in self-report measures. (12 2005). DOI:http://dx.doi.org/10.1002/nur.20100

[9] Jeffrey L Goodie, Kevin T Larkin, and Scott Schauss. 2000. Validation of Polar heart rate monitor for assessing heart rate during physical and mental stress. *Journal of Psychophysiology* 14, 3 (2000), 159.

[10] B L Heitmann and L Lissner. 1995. Dietary underreporting by obese individuals–is it specific or non-specific? *BMJ (Clinical research ed.)* 311, 7011 (10 1995), 986–9. DOI: http://dx.doi.org/10.1136/bmj.311.7011.986

[11] Jonathan Lester, Tanzeem Choudhury, Nicky Kern, Gaetano Borriello, and Blake Hannaford. 2005. A hybrid discriminative/generative approach for modeling human activities. In *roceedings of the international joint conference on Artificial intelligence (IJCAI)*.

[12] Robert Likamwa, Bodhi Priyantha, Matthai Philipose, Lin Zhong, and Paramvir Bahl. 2013. Energy characterization and optimization of image sensing toward continuous mobile vision. In *MobiSys 2013 - Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services*. 69–81. DOI: http://dx.doi.org/10.1145/2462456.2464448

[13] Hong Lu, Jun Yang, Zhigang Liu, Nicholas D. Lane, Tanzeem Choudhury, and Andrew T. Campbell. 2010. The Jigsaw continuous sensing engine for mobile phone applications. In *SenSys 2010 - Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*. 71–84. DOI: http://dx.doi.org/10.1145/1869983.1869992

[14] NIDDK. 2019. https://www.niddk.nih.gov/health-information/weight-management/adult-overweight-obesity/health-risks, Last accessed 2019-09-20. (2019).

[15] Sasank Reddy, Andrew Parker, Josh Hyman, Jeff Burke, Deborah Estrin, and Mark Hansen. 2007. Image browsing, processing, and clustering for participatory sensing: Lessons from a DietSense prototype. In *Proceedings of the 4th Workshop on Embedded Networked Sensors, EmNets 2007*. 13–17. DOI: http://dx.doi.org/10.1145/1278972.1278975

[16] Peregrin Spielholz, Barbara Silverstein, Michael Morgan, Harvey Checkoway, and Joel Kaufman. 2001. Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors. *Ergonomics* 44, 6 (5 2001), 588–613. DOI: http://dx.doi.org/10.1080/00140130118050

[17] Edison Thomaz. 2013. Practical food journaling. In *UbiComp 2013 Adjunct - Adjunct Publication of the 2013 ACM Conference on Ubiquitous Computing*. 355–360. DOI: http://dx.doi.org/10.1145/2494091.2501089

[18] Edison Thomaz, Irfan Essa, and Gregory D. Abowd. 2015. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In *UbiComp 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery, Inc, 1029–1040. DOI: http://dx.doi.org/10.1145/2750858.2807545