Activiome: A System for Annotating First-Person Photos and Multimodal Activity Sensor Data

Edison Thomaz
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, TX USA
ethomaz@utexas.edu

***Abstract*—Recognizing human activities in everyday environments remains a significant challenge in ubiquitous computing research. In this paper, we present Activiome, an end-to-end system designed to facilitate label annotation for activity recognition in real-world settings. It captures and processes in-situ multimodal sensor data and associated ground truth labels in the form of first-person photos. Participants can annotate their own data through a private web application interface. In addition to the Activiome system, which we are making available to the research community, our contributions include a detailed account of challenges and opportunities we encountered while developing and using the system over the last 5 years; this includes an evaluation of the Activiome mobile application's battery performance, a discussion of human concerns such as privacy, and a survey-based examination of comfort and perception issues affecting everyday usage of the system. We believe Activiome represents a valuable contribution to the community and its ongoing efforts to scale up activity recognition research beyond the walls of the laboratory.**

I. Introduction

Despite remarkable advances in sensors, mobile computing technologies and machine learning methods over the last decade, building systems that recognize complex human activities and behaviors in everyday environments remains a significant challenge. The problem stems in large part from the difficulty in obtaining reliable ground truth measures in naturalistic environments, which are required to train activity classifiers [1]. Directly observing individuals is considered to be the best method for compiling a log of activities performed in-the-wild, but it is not practical and could, in principle, alter the person's natural behavior. Prompting individuals to self-report daily activities with a journal or logbook has been a commonly-used strategy for estimating ground truth data in naturalistic settings [2]. Popular self-report techniques include the Experience Sampling Method (ESM) [3] and the Day Reconstruction Method (DRM) [4]. Unfortunately, these approaches can be highly disruptive, susceptible to biases, and difficult to sustain in the long term.

A new promising method for experience capture involves instrumenting individuals with wearable cameras. First-person perspective photographs depict individuals performing everyday activities and are rich in contextual detail. These characteristics make first-person photos well-suited as a source of ground truth. While obtaining first person images with wearable cameras can be done automatically, annotating thousands of images with ground truth labels, and associating them with multimodal sensor data streams is a significant challenge and burden. To facilitate this process, we developed a system called Activiome. It offers an end-to-end solution for capturing and processing in-situ sensor data and associated ground truth labels, which are obtained by reviewing first-person photos. It is comprised of three components: a mobile phone application, a web application, and a web back-end infrastructure. To this day, Activiome has contributed to research efforts totaling more than 100 participants spread across 6 in-the-wild studies; the majority of these studies focused on recognizing health behaviors such as eating and activities of daily living. As part of these experiments, more than 70,000 first-person images have been collected, reviewed and annotated.

In additional to Activiome, which we are making available to the research community, our contributions also include a detailed account of challenges and opportunities we encountered while developing and using the system. This analysis is highlighted with an evaluation of the Activiome mobile application's battery performance, a discussion of human concerns such as privacy, and a survey-based examination of comfort and perception issues affecting everyday usage of the system.

II. Related Work

Over the years, a variety of *capture and access* technologies aimed at the acquisition and visualization of human experiences have been developed [5]. Most of these systems were created for individual use targeting specific applications, such as helping people understand factors that influence their sleep environment [6] or emotional state [7]. One of the most ambitious efforts in this space was MyLifeBits [8], which explored whether all media reflecting a person's life, such as audio, video, documents, emails, phone calls, etc. could be captured, stored and made accessible for future use. Although these systems captured, analyzed and visualized sensor data, they were not designed as general purpose platforms for activity recognition.

Numerous research-oriented platforms have been developed to facilitate capture of experience samples. Froehlich et al. built MyExperience to enable studies of mobile technology usage and evaluation [9]. The system collected objective and subjective data on mobile devices, such as statistics on device usage, and featured context-triggered experience sampling.

AndWellness collected passive mobile phone sensor data and active user experience samples for health and behavior assessment [10]. Ramanathan et al. proposed ohmage, which offered similar functionality to AndWellness and was also motivated by opportunities in health research [11]. More recently, Xiong et al. built Sensus, a system designed specifically for mobile crowdsourcing with support for hardware and software sensors, automatic deployment of sensor-triggered surveys, and interface with a mobile platform [12].

The key difference between Activiome and systems like MyExperience, AndWellness and Sensus is that instead of relying on traditional experience sampling techniques for obtaining a measure of ground truth, our approach centers on the use of passively-captured first-person photos. Byrne et al. found these types of photographs to be particularly well-suited for task observations since it does not intrude into people's environment [13]. Indeed, this method has been extensively used in a variety of research applications [14]. For instance, Kelly et al. used it to evaluate travel-related patterns [15]; Image-Diet Day used automatically captured first-person photos in the context of dietary monitoring [16]; and Marcu et al. showed the value of such photos to support children with autism [17]. Our aim with Activiome is to provide a technological foundation for facilitating these types of research efforts.

III. Design Requirements

Built specifically for activity recognition research, Activiome was implemented around four design requirements; these were determined based on our experience with sensor data collection and annotation in real-world settings:

Real-Time Data Processing: There are many applications where it is desirable to react to incoming sensor data. For example, Just-in-Time Adaptive Interventions (JITAI) is an approach that targets problematic smoking and eating behaviors in the moment, and require the detection of activities in real-time [18]. An additional benefit of real-time data processing is being able to instantly identify if and when data collection problems occur, minimizing data loss.

Integration with Wearables: Leveraging off-the-shelf devices with embedded sensors devices is compelling because it facilitates long-term data collection with large populations, and in real-world settings. This contrasts with methods based on specialized sensors that are typically not as appealing to study participants and necessitate application-specific implementation and integration.

Low Annotation Effort: User burden is one of the most challenging aspects of gathering annotations in everyday settings. Although eliminating the annotation step altogether is not feasible, our goal was to minimize the time and effort involved in the labeling task as much as possible.

Privacy Protection: First-person photographs depict individuals performing everyday activities with unparalleled richness in detail and objectivity when compared to self-report methods. However, this method can pose a privacy threat to study participants. Therefore, it is critical to design an annotation workflow where only participants are allowed to review and label their own data.

Fig. 1. The Activiome on-body sensors: a mobile phone sitting on a lanyard around the neck captures first-person photos, GPS location, inertial data from the phone's sensors and short audio clips. Accelerometer and gyroscope data from two Movesense wearables are also collected by the Activiome mobile app.

IV. System Components

The Activiome system is made up of 3 high-level components: a mobile phone application, a web application, and a web back-end infrastructure. In addition to collecting multi-modal sensor data, the mobile application also integrates with the Movesense wearable device [1].

All sensor data and first-person photos acquired by the Activiome mobile phone application are uploaded in real-time to the Activiome web back-end infrastructure, which is organized around a web server and database. On top of the back-end infrastructure lies the web application, which researchers and individuals (e.g., study participants) use to review and annotate photos and sensor data.

A. Mobile Phone and Application

The mobile application is a key element of the system, capturing and aggregating multi-modal sensor data and first-person photos. The Activiome mobile app was implemented on the iOS platform at first, and we plan to port to the Android operating system in the future. Since the mobile phone running the app is programmed to function as a wearable camera, it must be worn continuously and thus requires a dedicated device. To minimize the impact of this requirement, we designed the app such that it does not need a top-of-the-line device; this allows older phones, such as the iPhone 4s running iOS version 7.0, to be re-purposed for this application.

[1]https://www.movesense.com

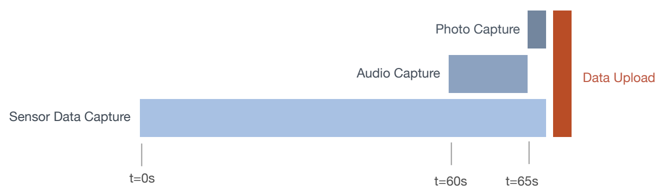

Fig. 2. The data acquisition cycle of the Activiome mobile app. In this example, the cycle is set to 60 seconds. After 60 seconds, the app first captures 5 seconds of audio and then takes a picture. Inertial sensor data is captured throughout the entire cycle. At the completion of data acquisition, the data is packaged as a HTTP POST request and uploaded to the Activiome server. This cycle repeats until the application quits.

In terms of form factor and placement, we have had good experience with a setup where the phone is worn around the neck on a lanyard with the rear-facing camera pointing forward. To extend the field-of-view of photographs, we usually instrument the Activiome phones with a wide-angle lens that attach to the phone case.

1) Sensing Modalities and Data Acquisition: The mobile application captures photos, audio, inertial sensing data and location following the data acquisition cycle illustrated in Figure 2. During the cycle, sensor data is captured at different rates and at different times. At the end of the cycle, all aggregated data is uploaded to the server as one HTTP POST request in the multipart/form-data format. For a six-hour data collection period, the amount of data uploaded is in the range of 150MB.

First-Person Photo: When the mobile application is running, it takes photos with the rear-facing camera at an user-configurable regular interval, which can be every 30 seconds or longer. This interval determines the duration of the data acquisition cycle runs. The photos are resized to 320x480 pixels and compressed to 0.5 quality (JPEG). A typical photo is around 15kb to 25kb in size. When GPS location is on, the photos are also geo-tagged with latitude and longitude.

Audio: An audio clip of either 5 or 10 seconds is recorded immediately prior to when first-person photos are shot. The audio is recorded in the MPEG4 autio format at a sample rate of 44.1Khz.

GPS: When enabled, location metadata is obtained with the iOS CoreLocation framework; latitude and longitude coordinates are set to be accurate to the nearest kilometer. The most recently updated coordinates are retrieved at the end of each cycle (i.e., right before the data is uploaded to the Activiome server).

Inertial: Accelerometer and gyroscope data are captured from different sources, (1) the iPhone that is running the Activiome mobile application, and (2) up to two Movesense sensors. In both cases, inertial data collection begins immediately after the application launches and takes place uninterruptedly even while the audio clip is being recorded. We have obtained inertial sensor data on the iPhone and with Movesense at 26Hz and 52Hz sample rates, respectively. To ease data collection

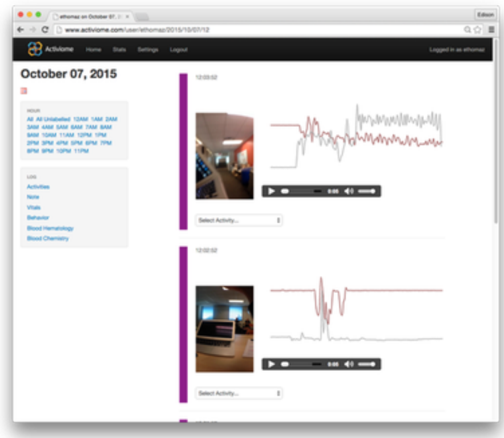

Fig. 3. The Activiome main screen. Once participants log in, they are shown a detailed list of recorded activities for the most recent hour

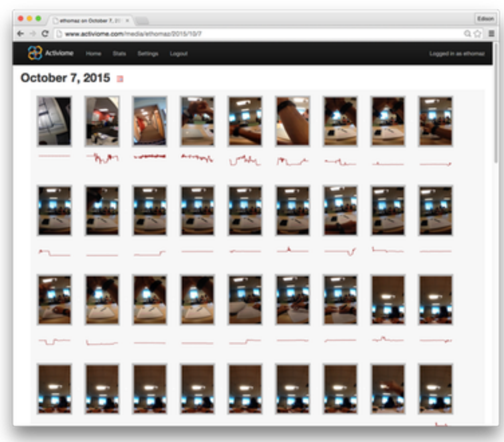

Fig. 4. The Activiome mosaic screen. To facilitate annotating large numbers of images, I created a photo view that shows thumbnails of the captured first-person point-of-view images and makes it easy to select and label multiple images at a time.

in lab studies, inertial data is also saved locally on the phone as flat files.

B. Web Application Interface and Annotation

Individuals interact with their data using the Activiome web application. Once individuals log in to their accounts, they see a list of activity entries for the day. Each entry on the web app interface includes a first-person photo, the audio clip, a graphical representation of the sensor data collected, and a drop-down menu with a pre-defined activity list, as shown in Figure 3.

A new activity entry gets created and is available for annotation at the end of each data acquisition cycle, which can be as short as 30-seconds. While a short acquisition cycle increases the temporal resolution of the data, as more activity entries get created, the annotation effort also increases. Each activity entry can be annotated individually or, to aid the ground truth labeling process of many entries at a time, the Activiome web application also offers a mosaic view, where thumbnails of all first-person photos taken on a given day are shown together (Figure 4). Using this view, participants can select multiple photos at a time using the keyboard or mouse and annotate their corresponding activity entries at once, reducing the time required for annotation significantly.

C. Server and Database

The backend infrastructure of Activiome was designed around well-established web technologies; it centers around a set of PHP scripts and a MySQL database. A PHP script is called by the mobile application with all the collected data in the HTTP POST request, and proceeds to parse and validate it. A new entry is created on the database and is populated with the sensor data, metadata (i.e., geo-location and timestamps), and links to the audio and image files. The audio clips and photographs are saved as flat files.

V. Use Cases

The case studies below illustrate how the combination of sensor data capture and first-person photo annotation provides a powerful foundation for activity recognition research in real-world settings.

A. Automated Dietary Monitoring

One of the studies where the Activiome system played a major supporting role focused on automated dietary monitoring. The research question addressed in the experiment was: “Is it possible to identify if a person is having a meal such as breakfast or lunch based on arm gestures captured with a smartwatch with inertial sensing capabilities?”. To train a food intake gesture classifier, data was collected in a lab setting with 20 participants; once the model was built, it was evaluated with more than 400 hours of sensor data captured in real-world settings over several days. The first-person images recorded by the Activiome mobile app and the web annotation interface were used to compile a ground-truth measure of when eating activities occurred.

B. Activity Recognition from First-Person Images

Recent advances in object detection with convolution neural networks has fueled interest in the application of computer vision techniques towards recognizing everyday activities. In a recent study, Activiome was used in the capture and compilation of perhaps the largest annotated dataset of first-person images. Over 40,000 photos were recorded over a six month period using the Activiome mobile application; each photo was labeled as belonging to one of 19 activity classes with the Activiome web application. This work demonstrated the feasibility of recognizing a large array of human activities from first-person photographs and temporal metadata.

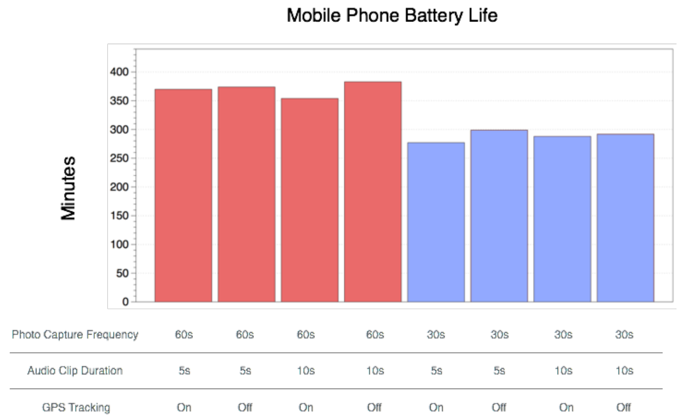

Fig. 5. We performed an analysis of battery consumption in naturalistic environments under different configuration and learned that the frequency at which first-person photos is captured is the primary factor determining battery usage.

VI. Challenges & Opportunities

Over the last 3 years, Activiome has served as a foundation for several human behavior analysis research efforts. In this section we describe challenges we faced, and opportunities that emerged from our experiences of deploying Activiome with participants in real-world settings.

A. Battery Life & Performance

Considering the workload of the Activiome mobile application, it is not surprising that battery life is a serious concern when it comes to recording first-person photos, audio clips and other forms of sensor data. We performed an analysis of battery consumption in naturalistic environments under different configuration and learned that the frequency at which first-person photos is captured, tested at 60 and 30 seconds, is the primary factor determining battery usage (Figure 5). On average, the phone battery lasted for 370 and 289 minutes for the 60 and 30-second configurations, respectively. As a result, an external battery (5200mAh capacity) plugged to the phone proved to be a necessity for in-the-wild studies lasting longer than 6 hours.

From our performance analysis, we observed that audio clip recording and GPS tracking did not have a meaningful impact on battery life. Very low battery consumption was also noted with Movesense, which is powered by a 3V lithium coin cell battery, and is engineered specifically for low-power use.

B. Data Resolution

Collecting first-person photos and sensor streams every 30 seconds makes it possible to record individuals’ life activities throughout the day. For many applications, such as inferring if individuals travel to work by car or public transportation, this level of resolution is enough. However, there are many compelling scenarios where increased resolution is required. In health research, for instance, it is often desirable to spot eating and smoking gestures. These types of gestures are very short in duration, in the range of 2 to 5 seconds, so first-person photographs taken every 30 seconds rarely capture them.

The obvious solution to this problem is to increase the frequency at which photos are taken, or record video instead of photographs. Unfortunately, mobile phones today can only capture photos or videos continuously for a small portion of the day before the battery is completely drained. To address this limitation, an external battery can be paired with the phone, at the cost of additional weight and inconvenience. Another challenge with data collection at a higher frequency involves the annotation task; having to label more photos or review videos makes the annotation task even more onerous and time consuming.

C. Multiple Labels

Using the web application interface, photos and the underlying sensor data can be tagged with only one activity label at a time. The decision to limit this association to a one-to-one mapping was made to simplify the annotation process in line with the stated "Low Annotation Effort" design goal. If multiple labels could be chosen per photo, individuals would have to spend more time examining each one, which would certainly increase burden in the form of cognitive load and time spent on the task.

From the point of view of accuracy, having to choose only one label is clearly a limiting factor as people constantly multitask throughout the day. We are investigating techniques that would allow annotators to select multiple labels quickly and without a heavy cognitive penalty. In our experience, however, we found that this limitation could be mitigated by instructing annotators to choose an activity label that represented the *primary* activity depicted in the photo. This approach was only possible because the annotators were typically the study participants themselves, and they could recall their activities from memory.

D. Privacy Considerations

First-person images offer many opportunities in activity recognition research but they also pose significant ethical challenges, particularly with regards to privacy. Obtaining informed consent from third-parties for capturing and reviewing first-person images remains an open research problem. In previous work, Nguyen et al. discussed acceptable boundaries and hurdles for usage of wearable cameras in public settings [19], and others have proposed frameworks and techniques for minimizing these challenges [20], [21].

In our experience, we have found that a practical compromise can be reached without explicit third-party consent. The solution, approved by our study review board and subsequently incorporated into the design of Activiome, is to have study participants only review and annotate their own photographs, which is achieved through individualized login accounts. The rationale for this approach stems from the observation that in principle, first-person photos capture moments that the individuals carrying the mobile device already experienced. Therefore, any photo that bystanders wished it had not been captured are merely a representation of a memory held by the individual carrying the mobile device. This holds true as long the photo is kept safe and never shared with anyone else. As an additional layer of privacy, individuals do not need to provide personally identifiable information when creating an Activiome account. Therefore, even if a server administrator were to gain direct access to user data, it would not be obvious who it belongs to.

E. Comfort and Perception

Considering the instrumentation required by Activiome, a phone sitting on a lanyard around the neck; a Movesense wearable on each wrist; and an external battery (only required for day-long user studies), we were interested in participants' perception of the system after having used it for a period of time in real-world settings. Our inquiry focused on the mobile phone and its unconventional placement.

We asked ten participants who had just completed a 3-day study to answer a short questionnaire structured around a five-point Likert scale with two Likert items: "Wearing the camera was comfortable", and "I was self-conscious when wearing the camera in public". Possible answers were *"Strongly Disagree (SD)"*, *"Disagree (D)"*, *"Neither agree nor disagree (NAND)"*, *"Agree (A)"*, and *"Strongly Agree (SA)"*. We chose to describe the mobile phone as a camera in the statements because we informally identified that the picture-taking capability of the device was what concerned participants the most.

The distribution of answers for both items can be see in Figure 6. Half of the participants agreed that wearing the phone-camera was comfortable, while 4 neither agreed nor disagreed. A few participants wrote down notes next to their ratings; one participant complained that *"[the phone] moved around a lot in its holder"*. Another claimed having *"issues with the phone slipping out of the case"*. More concerning was the feedback from one participant who said that the *"weight [of the phone] made my neck a bit tired"*. In fact, the only withdrawal we have ever had in a study with the Activiome system was for this exact reason: neck pain due to the weight of the mobile phone. Feelings were more mixed about whether usage of the device caused self-consciousness; most neither agreed nor disagreed. A quote from one participant who was neutral in his Likert-scale response was *"sometimes, like people would ask me why I'm 'wearing' my phone? :-)"*.

In light of these findings, we are exploring different form factors for the Activiome on-body sensors, and the mobile phone in particular. A smaller and more lightweight design approach will likely result in comfort improvements and reduce the self-consciousness effect caused by the presence of a phone around the neck.

VII. Conclusion

The quantity, complexity, and variability of human behaviors makes the development of activity recognition systems a challenging undertaking. A fundamental difficulty is the acquisition of ground truth labels for training classifiers.

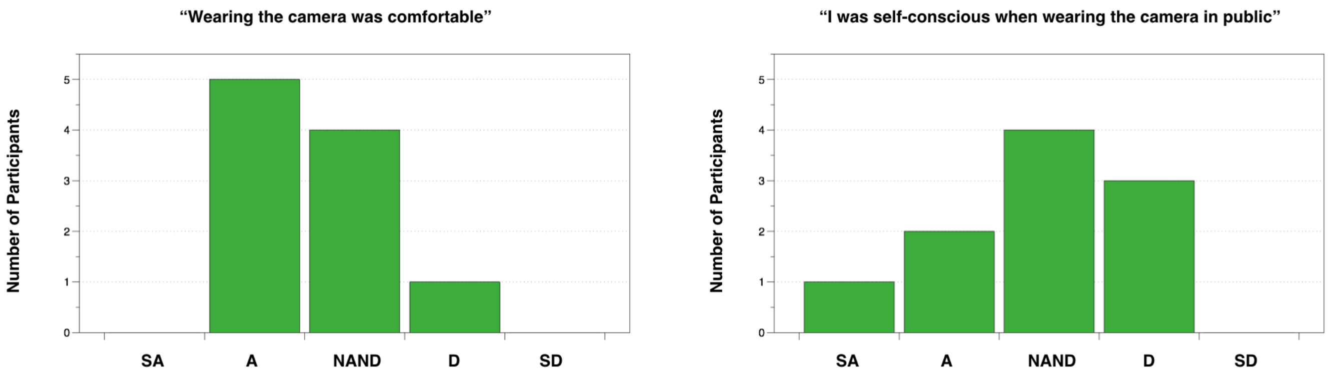

Fig. 6. Five-point Likert scale ratings for 10 participants on two Likert items: "Wearing the camera was comfortable", and "I was self-conscious when wearing the camera in public". Possible answers were Strongly Disagree (SD), Disagree (D), Neither agree nor disagree (NAND), Agree (A), and Strongly Agree (SA).

Activiome facilitates this task by capturing multimodal sensor data and offering tools for mapping said sensor data to annotations. Although there are clear opportunites for improvement, we believe Activiome represents a valuable contribution to the community and its ongoing efforts to scale up activity recognition research beyond the walls of the laboratory.

References

[1] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, p. 33, 2014.

[2] S. S. Intille, L. Bao, E. M. Tapia, and J. Rondoni, "Acquiring in situ training data for context-aware ubiquitous computing applications," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2004, pp. 1–8.

[3] R. Larson and M. Csikszentmihalyi, "The experience sampling method in flow and the foundations of positive psychology (pp. 21–34)," 2014.

[4] D. Kahneman, A. B. Krueger, D. A. Schkade, N. Schwarz, and A. A. Stone, "A survey method for characterizing daily life experience: The day reconstruction method," *Science*, vol. 306, no. 5702, pp. 1776–1780, 2004.

[5] K. N. Truong, G. R. Hayes *et al.*, "Ubiquitous computing for capture and access," *Foundations and Trends® in Human–Computer Interaction*, vol. 2, no. 2, pp. 95–171, 2009.

[6] M. Kay, E. K. Choe, J. Shepherd, B. Greenstein, N. Watson, S. Consolvo, and J. A. Kientz, "Lullaby: a capture & access system for understanding the sleep environment," in *Proceedings of the 2012 ACM conference on ubiquitous computing*. ACM, 2012, pp. 226–234.

[7] D. McDuff, A. Karlson, A. Kapoor, A. Roseway, and M. Czerwinski, "Affectaura: an intelligent system for emotional memory," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012, pp. 849–858.

[8] J. Gemmell, G. Bell, R. Lueder, S. Drucker, and C. Wong, "Mylifebits: fulfilling the memex vision," in *Proceedings of the tenth ACM international conference on Multimedia*. ACM, 2002, pp. 235–238.

[9] J. Froehlich, M. Y. Chen, S. Consolvo, B. Harrison, and J. A. Landay, "Myexperience: a system for in situ tracing and capturing of user feedback on mobile phones," in *Proceedings of the 5th international conference on Mobile systems, applications and services*. ACM, 2007, pp. 57–70.

[10] J. Hicks, N. Ramanathan, D. Kim, M. Monibi, J. Selsky, M. Hansen, and D. Estrin, "Andwellness: an open mobile system for activity and experience sampling," in *Wireless Health 2010*. ACM, 2010, pp. 34–43.

[11] N. Ramanathan, F. Alquaddoomi, H. Falaki, D. George, C.-K. Hsieh, J. Jenkins, C. Ketcham, B. Longstaff, J. Ooms, J. Selsky *et al.*, "Ohmage: an open mobile system for activity and experience sampling," in *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*. IEEE, 2012, pp. 203–204.

[12] H. Xiong, Y. Huang, L. E. Barnes, and M. S. Gerber, "Sensus: a cross-platform, general-purpose system for mobile crowdsensing in human-subject studies," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 415–426.

[13] D. Byrne, A. R. Doherty, G. J. Jones, A. F. Smeaton, S. Kumpulainen, and K. Järvelin, "The sensecam as a tool for task observation," in *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 2*. BCS Learning & Development Ltd., 2008, pp. 19–22.

[14] A. R. Doherty, S. E. Hodges, A. C. King, A. F. Smeaton, E. Berry, C. J. Moulin, S. Lindley, P. Kelly, and C. Foster, "Wearable cameras in health: the state of the art and future possibilities," *American journal of preventive medicine*, vol. 44, no. 3, pp. 320–323, 2013.

[15] P. Kelly, A. Doherty, E. Berry, S. Hodges, A. M. Batterham, and C. Foster, "Can we use digital life-log images to investigate active and sedentary travel behaviour? results from a pilot study," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 8, no. 1, p. 44, 2011.

[16] L. Arab, D. Estrin, D. H. Kim, J. Burke, and J. Goldman, "Feasibility testing of an automated image-capture method to aid dietary recall," *European journal of clinical nutrition*, vol. 65, no. 10, p. 1156, 2011.

[17] G. Marcu, A. K. Dey, and S. Kiesler, "Parent-driven use of wearable cameras for autism support: a field study with families," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 401–410.

[18] I. Nahum-Shani, S. N. Smith, B. J. Spring, L. M. Collins, K. Witkiewitz, A. Tewari, and S. A. Murphy, "Just-in-time adaptive interventions (jitais) in mobile health: key components and design principles for ongoing health behavior support," *Annals of Behavioral Medicine*, vol. 52, no. 6, pp. 446–462, 2017.

[19] D. H. Nguyen, G. Marcu, G. R. Hayes, K. N. Truong, J. Scott, M. Langheinrich, and C. Roduner, "Encountering sensecam: personal recording technologies in everyday life," in *Proceedings of the 11th international conference on Ubiquitous computing*. ACM, 2009, pp. 165–174.

[20] P. Kelly, S. J. Marshall, H. Badland, J. Kerr, M. Oliver, A. R. Doherty, and C. Foster, "An ethical framework for automated, wearable cameras in health behavior research," *American journal of preventive medicine*, vol. 44, no. 3, pp. 314–319, 2013.

[21] 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 *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 2013, pp. 739–748.