Association for Information Systems
AIS Electronic Library (AISeL)

Research Papers | ECIS 2018 Proceedings

11-28-2018

Using Deep Learning and 360 Video to Detect Eating Behavior for User Assistance Systems

Philipp Vincent Rouast
The University of Newcastle, philipp.rouast@uon.edu.au

Marc Adam
University of Newcastle, marc.adam@newcastle.edu.au

Tracy Burrows
The University of Newcastle, tracy.burrows@newcastle.edu.au

Raymond Chiong
The University of Newcastle, raymond.chiong@newcastle.edu.au

Follow this and additional works at: https://aisel.aisnet.org/ecis2018_rp

Recommended Citation

Rouast, Philipp Vincent; Adam, Marc; Burrows, Tracy; and Chiong, Raymond, "Using Deep Learning and 360 Video to Detect Eating Behavior for User Assistance Systems" (2018). *Research Papers*. 101.
https://aisel.aisnet.org/ecis2018_rp/101

This material is brought to you by the ECIS 2018 Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in Research Papers by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

USING DEEP LEARNING AND 360 VIDEO TO DETECT EATING BEHAVIOR FOR USER ASSISTANCE SYSTEMS

Research in Progress

Rouast, Philipp V., The University of Newcastle, Callaghan, Australia, philipp.rouast@uon.edu.au

Adam, Marc T.P., The University of Newcastle, Callaghan, Australia, marc.adam@newcastle.edu.au

Burrows, Tracy, The University of Newcastle, Callaghan, Australia, tracy.burrows@newcastle.edu.au

Chiong, Raymond, The University of Newcastle, Callaghan, Australia, raymond.chiong@newcastle.edu.au

Rollo, Megan, The University of Newcastle, Callaghan, Australia, megan.rollo@newcastle.edu.au

Abstract

The rising prevalence of non-communicable diseases calls for more sophisticated approaches to support individuals in engaging in healthy lifestyle behaviors, particularly in terms of their dietary intake. Building on recent advances in information technology, user assistance systems hold the potential of combining active and passive data collection methods to monitor dietary intake and, subsequently, to support individuals in making better decisions about their diet. In this paper, we review the state-of-the-art in active and passive dietary monitoring along with the issues being faced. Building on this groundwork, we propose a research framework for user assistance systems that combine active and passive methods with three distinct levels of assistance. Finally, we outline a proof-of-concept study using video obtained from a 360-degree camera to automatically detect eating behavior from video data as a source of passive dietary monitoring for decision support.

Keywords: Decision support, user assistance systems, dietary monitoring, deep learning.

1 Introduction

Non-communicable diseases such as cardiovascular disease, cancer, and diabetes are the leading cause of death globally (WHO, 2017). Major contributing factors include common lifestyle choices such as unhealthy diet and insufficient physical activity. Obesity continues to rise as every third U.S. adult is now obese (Hales et al., 2017). Although information regarding better nutrition is widely available via information systems (IS), current studies indicate that the intended effect on eating behavior is limited. For instance, in Australia less than 4% of adults eat sufficient amounts of vegetables per day, whereas one third of their calories are from "junk" foods (Australian Bureau of Statistics, 2016). Interventions are lacking targeted guidance for specific goals such as increased intake of dietary fiber. The desired behavior change requires us to address not only capability by providing information to users, but also opportunity and motivation (Michie et al., 2011). IS designed to achieve effective behavior change need to consider these factors to deliver targeted assistance, counteracting the tendencies introduced by the population's shift to a largely sedentary lifestyle.

Dietary monitoring lays the foundation for personalized behavior change interventions by recording *what*, *how much*, and *when* foods and drinks are consumed. Active methods, the most-widely used approach in practice, achieve this by relying on self-report by the user. However, dietary monitoring using these methods can be highly time-consuming and burdensome. It consequently faces a lower adoption than the related energy expenditure monitoring, which largely relies on passive sensors (Krebs and Duncan, 2015). While an increasing body of research is emerging on passive methods of dietary monitoring (Vu et al., 2017), it has yet to produce viable standalone artefacts for mainstream use. Modern computer algorithms based on machine learning, especially deep learning (LeCun et al., 2015), are expected to improve on this.

In this paper, we advocate the use of passive methods of dietary monitoring in combination with active methods to provide user assistance on improving dietary intake for improved health outcomes. Passive methods enabled by sensor data and deep learning techniques play a key role by gathering real-time information in an unobtrusive manner, as this information may assist the user with the use of active methods and hence improve the overall quality of dietary assessment. Considering the goal of behavior change, this allows us to address the factors of opportunity and motivation rather than merely providing the user with information. Based on the taxonomy by Morana et al. (2017), we develop a framework for a system providing user assistance in the context of diet and nutrition considering the following components: (i) improvement of dietary monitoring by suggestive prompting, pre-completion, and review of data entry, (ii) informing nutrition decisions by leveraging dietary monitoring for informational assistance, and (iii) situation-aware invocations for proactive assistance based on real-time information. Finally, we report our progress on a proof-of-concept study in which we assess the feasibility of passively obtaining data for such systems. For this purpose, we use videos of eating occasions from a 360-degree camera and deep learning techniques.

2 Background and Literature Review

Dietary assessment pursuits the goal of obtaining unbiased data on typical food intake. By contrast, dietary monitoring aims at changing behavior or reinforcing positive behavior. Individual and situation-aware assistance helps to achieve these goals, thus requiring accurate and reliable data on food intake and eating habits. In the following sub-sections, we review active and passive methods to acquire such data. We determine which methods are used in practice today, as well as their drawbacks and future potential. We also give a short primer on user assistance systems in the context of dietary monitoring, what purposes these systems serve, and challenges in their establishment.

2.1 Active Methods of Dietary Assessment and Monitoring

Dietary assessment can be the basis for guidance by dietitians or monitoring systems; it is traditionally conducted in the form of food records, recalls, or specialized questionnaires (Block, 1982). These methods require subjects to record all individual consumed foods along with corresponding estimated or measured weights. The recall and questionnaire methods rely on memory, as intake amounts and frequencies are logged at intervals such as 24 hours or 7 days. With growing availability of consumer-grade cameras, images taken of meals before and after eating have also been used widely to report food intake to dietitians and reduce user burden (Ashman et al., 2017). In the age of the smartphone and portable devices, traditional methods of keeping records on paper have been replaced by dedicated mobile and web applications connected to large food databases that somewhat simplify the process (Darby et al., 2016). Many available applications now serve the purpose of active self-monitoring.

Active methods of dietary assessment and monitoring, such as the ones discussed above, have enabled dietitians' work for decades and made countless contributions to research possible. However, some areas of improvement remain. For one, they are perceived as burdensome by the user (Turner-McGrievy et al., 2013), hence there is potential to employ IS to reduce the perceived effort of keeping food records. Furthermore, manual reports have been found to be inaccurate in many cases. Several studies suggest that individuals report expected intake instead of real intake (Lichtman et al., 1992; Westerterp and Goris, 2002), by failing to report some meals and snacks, incorrectly labelling foods, and misjudging portion sizes. To remedy this effect, sensor-based systems could support users in entering their meals. Objectively detected values could be used as suggestions to nudge users in the right direction. Research also suggests that individuals request more behavioral and interactive elements when monitoring dietary intake (Lee et al., 2017; Solbrig et al., 2017), to make the process more engaging (Burke et al., 2017). Especially for recall, active assessment and monitoring methods do not emphasize the temporal and behavioral aspects of food consumption. They are designed for manual review where corresponding feedback is delivered at a delay. Real-time information with accurate timestamps for eating occasions allows IS to provide instantaneous decision support to users concerning their eating behavior.

2.2 Passive Methods of Dietary Assessment and Monitoring

Approaches to passive dietary assessment take advantage of the characteristic properties of human food intake and mechanical digestion processes; this enables them to automatically gather information through a variety of sensors. They typically handle the tasks of eating behavior detection, food type classification, and volume or weight estimation (Vu et al., 2017).

Current research mainly focuses on visual, acoustic, and inertial means of collecting information. Some work has been done to automate the process of estimating calories from images of food using computer vision and machine learning. Such techniques may be applied to images taken with everyday cameras in active assessment (Zhu et al., 2010), or integrated in passive systems that automatically take such images (Sun et al., 2014). Visual signals have also been used in video acquired from stationary cameras to automatically detect chewing events. In this visual approach, food classification accuracies of 80%-90% can be achieved, albeit based on limited types of food (Vu et al., 2017). An approach based on deep learning reports a classification accuracy of 72% on 100 classes of food (Kawano and Yanai, 2014). Swallowing and chewing events as well as basic food types are shown to be detectable in acoustic approaches using wearable microphones (Amft et al., 2005). Similarly, the inertial approach detects food intake events using wrist-worn gyroscopes or accelerometers typically available in smartwatches (Dong et al., 2012). The acoustic and inertial approaches can achieve up to 85% of accuracy for swallowing detection, 90% for eating detection, and 94% for eating gesture detection (Vu et al., 2017). Further approaches explored in the literature involve physiological and piezoelectric signals, as well as fusions of different approaches (Vu et al., 2017).

In general, accuracies of food type classification and weight estimation present some of the main challenges for the passive methods. Although the goal of fully autonomous dietary assessment in real-life situations proves to be too difficult for the state-of-the-art, one can imagine useful applications of the information gathered by prototypes to improve existing active assessment and monitoring systems. At the same time, we can expect substantial improvements in accuracy and scope of passive systems as both sensors and machine learning research are progressing.

2.3 User Assistance Systems

In the age of the smartphone, it is technologically feasible to design IS that allow users to actively capture virtually all aspects of their diet. However, the practical feasibility of systems that solely rely on active capture methods is questionable, as they require a high level of user involvement. Hence, it does not come as a surprise that dietary monitoring systems building on active capture methods are often perceived as time-consuming and burdensome which in turn detrimentally affects the quality of dietary assessment. A promising context that has received increasing research attention in recent years are so-called *user assistance systems* (UAS). UAS are software components that enrich information systems aiming to assist users to perform their tasks better (Maedche et al., 2016). A common form of assistance provided in such UAS are guidance design features, that can be further classified along dimensions such as target, directivity, mode, and invocation (Morana et al., 2017). In this context, past failures like Microsoft's *Clippy* exemplify that UAS are not trivial to design, and that it is crucial to exhibit context-awareness. Considering dietary monitoring, data collected with passive capture methods (e.g., sound, video) paired with the advances in the field of artificial intelligence have the potential to provide such systems with information to base assistance and guidance features on.

3 Theoretical Foundations

3.1 Framework

To support individuals in making better decisions about their diet, we consider the components that influence human behavior: Capability is linked to an individual's knowledge and can be addressed by providing information such as dietary advice. This is the primary focus of traditional approaches and most systems available today. Motivation and opportunity refer to internal and external factors affecting an individual's decision making, respectively. An intervention aiming at changing behavior, such as dietary intake, should take all three components into account simultaneously (Atkins and Michie, 2015; Michie et al., 2011; Aljaroodi et al., 2017).

In terms of the design of IS, the *Fogg Behavior Model* (Fogg, 2009) postulates that for a system to be persuasive in influencing user behavior, it must address three factors: The user must be sufficiently *motivated* and have the *ability* to perform the behavior, while at the same time being *triggered* to do so. Evidently, existing dietary monitoring systems building on active methods don't achieve high enough levels of ability and motivation to keep users engaged long enough: While users are initially motivated by the hope of becoming healthier, findings from recent research suggest that they find methods purely based on active monitoring burdensome, and that a lack of assistance leads to fading motivation (Solbrig et al., 2017). In order to sustain this motivation, user assistance is needed that makes interaction with the system simpler by targeting users' ability (Oinas-Kukkonen and Harjumaa, 2009). It includes factors such as time, physical activity, and brain cycles (Fogg, 2009). An additional caveat of systems relying on active methods is a lack of information to effectively place triggers. Users of nutrition-related systems prefer to receive relevant notifications and reminders (Krebs and Duncan, 2015), but without context-awareness of the system, these may appear at inappropriate points in time.

In Figure 1, we propose a framework for a UAS in this setting, building on the taxonomy by Morana et al. (2017). We argue that combining passive methods for food intake monitoring alongside active methods can be effectively integrated into a UAS that contributes to facilitating behavior change for healthy nutrition. Informed by passively monitored data, user assistance features can address the problems previously identified with system persuasiveness. This, in turn, allows us to better address motivation and opportunity. Assistance should rely on interaction with the user and exhibit intelligence by being situation-aware. Artefacts with these capabilities have also been referred to as anticipating UAS, and are argued to be the logical next step in user assistance within IS (Maedche et al., 2016).

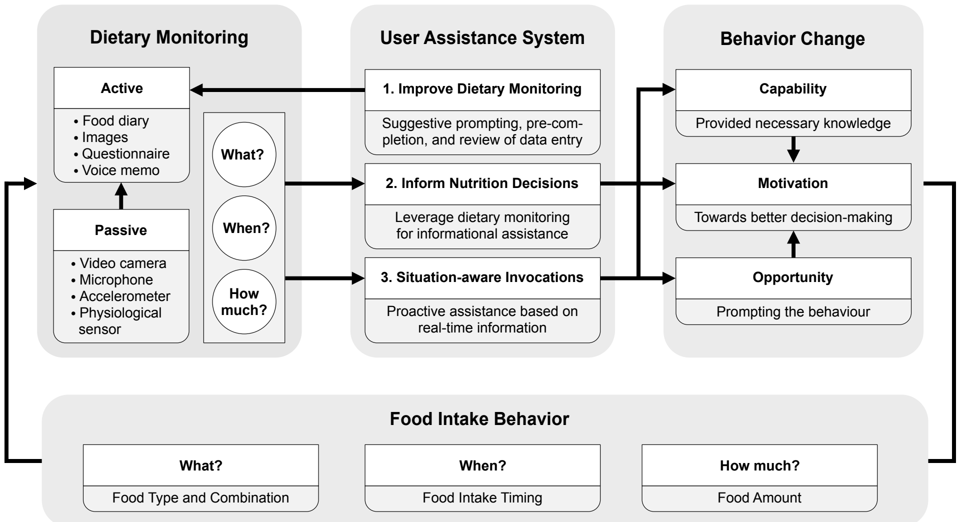

Figure 1. A user assistance systems framework for dietary monitoring.

In specifying the UAS, we follow the taxonomy of Morana et al. (2017) to describe the outlined guidance design features: Assistance should be provided both during eating occasions and when reviewing a diet.

Both informative and suggestive assistance should be provided, as detailed in the next section. Assistance can be invoked either directly by the user or intelligently based on the user's behavior; timing can occur concurrently or retrospectively to the corresponding user activity. The most obvious format of assistance is text-based. However, multimedia formats such as audio could also be explored. The assistance intention is to provide dietary recommendations, which aim to have some learning effect for the user. As the underlying machine learning models are trained in a supervised way from a database of examples labelled by experts, the delivered content type is adapted expert knowledge. The audience will mostly be novice users, but we anticipate that experts with a dietetics background will want to rigorously assess the performance of passive systems. Robust systems implementing this framework will incorporate proactive trust-building through accuracy and helpfulness of recommended assistance.

3.2 Levels of Assistance

As illustrated in Figure 1, the UAS integrates active and passive monitoring. The passive component is a key element, although the user does not directly interact with it. As it facilitates the gathering of real-time information, the system becomes situation-aware: The nature of the available data creates opportunities to complement the active component and to provide proactive user assistance. This is instrumental in enriching and improving the dietary assessment, lowering the burden of active monitoring on the user, and then targeting user capability, motivation, and opportunity to support behavior change. Three levels within the UAS work to achieve these goals:

1. *Improve the dietary monitoring* by suggestive prompting, pre-completion, and review of data entry

   During food intake, passive monitoring provides data from available sensors (e.g., video frames of eating occasion, accelerometer data, or audio) to the UAS. Automatic detection of eating behavior allows the system to infer the timing of individual eating occasions (the *when*). Firstly, this opportunity allows the system to remind the user that a meal should be logged – failing to record meals and snacks is a known contributor to under-reporting (Westerterp and Goris, 2002). Secondly, automatic classification of food type consumed during eating occasions (the *what*) and food volume estimation (the *how much*) have the potential to be used for pre-completion or assistance during active monitoring. Such assistance may help to reduce effort associated with manual entry. Available data can also help to point out errors in user entry if the estimates are far off. Progress in machine learning and sensor technology will lead to a gradual decrease in user effort and obtrusiveness associated with active entry, whilst increasing robustness and accuracy. Ultimately, the user's active contribution could be reduced to confirming values reported by the passive component.

2. *Inform nutrition decisions* by leveraging dietary monitoring for informational assistance

   Given an accurate account of dietary intake and associated habits based on active and passive monitoring, the UAS can dynamically derive personalized nutrition briefings (e.g., daily) to address the user's psychological *capability* and *motivation*. These have the function of (i) providing motivational support regarding recent data, projecting how well the user has done, and how soon specific goals can be achieved; (ii) the educational effect of pointing out how behavior changes contribute to user health; and (iii) recommendations for optimally adjusting diet and goals. Personalized delivery of supportive information takes the role of the advice given by a dietitian: It has potential to keep the user motivated (Solbrig et al., 2017) and improve the user's capability to realize a behavior change.

3. *Situation-aware invocations* for proactive assistance based on real-time information

   It is well known that goals are most likely to be achieved if they are tailored to the individual situation of the person. However, providing feedback is also important (Locke and Latham, 2002). Appropriately timed and relevant reminders allow the UAS to target *opportunity* and *motivation*. Since passive monitoring enriches the system with real-time information, the system can provide context-sensitive feedback. This feedback should be designed in a way that keeps the user motivated and engaged, reminding them of their health goals. It may include hints indicating that a target value is close to being reached (e.g., total amount of fiber consumed). Research shows that eating behavior with specific time signatures (e.g., night, weekend), is often linked to adverse health effects due to overeating (Vu et al., 2017). If such tendencies are detected, the user could be informed of this potentially unhealthy eating behavior and given information on how to improve. The rate of eating is also of interest, as it has been shown to have positive effect on obesity (Sasaki et al., 2003) – detection of such behavior could trigger similar hints reminding the user to slow down.

3.3 Deep Learning for Nutrition Monitoring

In passive nutrition monitoring, we require algorithms to handle complex real-time detection, prediction, and estimation tasks on multimodal sensor datasets. Driven by the availability of more computational power, large labelled datasets and advanced learning algorithms, the field of machine learning has received a lot of attention in the past years. Deep learning, the state-of-the-art in artificial neural networks, boasts unprecedented results in problems characterized by highly varying functions such as speech and object recognition (LeCun et al., 2015). The recent interest in deep learning primarily originated from the advances in computer vision research (Krizhevsky et al., 2012) using convolutional neural networks (CNNs) that are inspired by biological visual systems and designed specifically to learn from image data. An increasing amount of research and use in practice involves advanced recurrent neural networks (RNNs) (Hochreiter and Schmidhuber, 1997) that use a 'memory' component to master the intricacies of sequential data such as text, audio and video. As the capabilities of CNNs in object- and action recognition from visual and auditory sensor data are approaching human level, such architectures are also suitable for working with recordings of eating occasions. Within the UAS framework introduced in Section 3.2, we see the following application areas for deep learning:

1. *Detection of information regarding eating occasions from sensor data (level 1)*. Deep learning in the form of CNNs and RNNs could be applied to detect information regarding the *when* (e.g., detecting eating occasions, individual hand-to-mouth movements, chewing or swallowing from video, audio recordings, or inertial sensor data), the *what* (e.g., food type classification from image or video data), and the *how much* (e.g., food volume estimation from image or video data).
2. *Prediction of suitable timings for notifications and triggers (level 3)*. Given contextual information and usage patterns, interactive systems could use deep reinforcement learning to predict suitable points in time to deliver notifications and triggers (Christiano et al., 2017).

4 Research Methodology

Building on the framework introduced in the previous section, our research aims at assessing the feasibility of a surveillance video-based passive food intake monitoring system as data source for a UAS. This is an exemplary implementation of the first application area of deep learning proposed in Section 3.3. The approach is to use a standalone camera instead of requiring subjects to be wearing individual dedicated sensors. This approach is favored for its simplicity, unobtrusiveness, and cost-effectiveness, allowing the monitoring of multiple people in parallel by placing a 360-degree video camera in the center of a dining table as shown in Figure 2. A deep network is trained to learn from a database of labelled videos depicting eating occasions in a controlled eating environment. The study is comprised of three stages: Creation of a video database, training a model to detect basic eating behavior (targeting the *when*), and training a model to detect more advanced information (targeting the *what* and *how much*).

Stage 1. A sample of 100 healthy participants without food allergies will be recruited on campus using flyers and a social media campaign. At the estimated rate of 40 hand-to-mouth movements per eating occasions, this will provide enough examples to train and test a deep network. During the experiment, participants are provided with a standard meal consisting of different types of food that require fingers, knife and fork, and spoon to eat. Since they can be identified from video recordings, transparency about data use and protection of their privacy is given special attention. Participants receive an information statement when signing up for the experiment, and are asked to sign a consent form to participate. Inclusion of their data in a dataset for the larger research community and usage in publications are strictly opt-in. Otherwise, all data is stored securely on university servers and solely used during labelling, and to train and test our models. This data collection has been approved by the ethics committee at The University of Newcastle. For each session, four participants are asked to eat from a pre-prepared meal;

images of the meal are recorded before and after consumption. Participants are also asked to complete a post-experiment questionnaire which explores the factors that determine users' technology acceptance for this application. We expect the individual eating sessions to take no longer than 30 minutes. Subsequently, the videos are labelled by experts with information of interest: calories consumed, as well as hand-to-mouth movements with associated timestamps, used vessel, and food type.

Progress: Since confirmation of ethics approval, a pilot experiment of two sessions has been conducted; the recorded videos have been processed and labelled, amounting to a total of eight eating occasions.

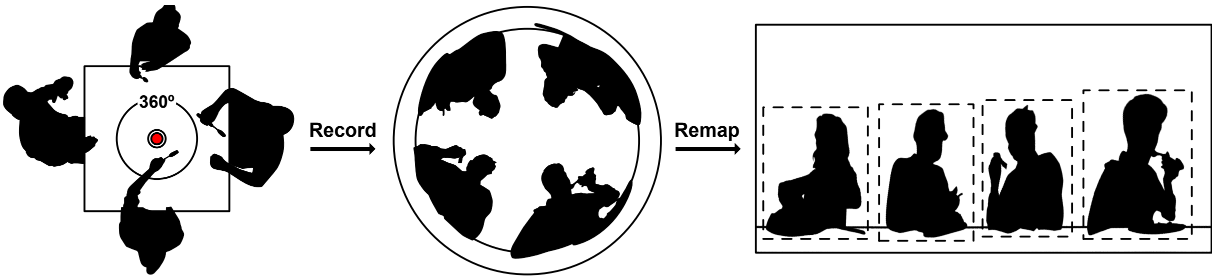

Figure 2. The video is recorded spherically and then remapped to an equirectangular representation for further segmentation and processing.

Stage 2. Initially, we consider the task of automatically detecting individual hand-to-mouth movements during eating occasions. These allow us to derive measures like total duration of an eating occasion, eating pace, number of bites consumed, as well as bite-to-bite intervals. For image-based classification, CNNs are the preferred choice in the literature, while RNNs aid in the modelling of sequential data. We will explore different configurations and combinations of both to learn from the sequential video data. As illustrated in Figure 2, pre-processing consists of mapping a spherical recording to equirectangular representation, and segmenting the individual subjects.

Progress: We used TensorFlow (Abadi et al., 2016) to train a six-layer CNN from scratch on our pilot dataset. We use four convolutional layers and two fully connected layers. This is a simplified version of the well-known AlexNet (Krizhevsky et al., 2012) architecture, a CNN designed for visual recognition. For each frame, the binary classification determines whether the subject is engaged in a hand-to-mouth movement or idle. Test data consists of one subject unknown to the trained network. Although we achieve 70% accuracy on the class-balanced test set, there is a considerable amount of overfitting attributable to the small amount of data. We expect significant improvements in accuracy with the larger dataset of 100 participants and architecture improvements such as regularization and sequence modelling using an RNN. Figure 3 illustrates example data the network sees during training and test.

Figure 3. Example frames from the pilot experiment

Stage 3. In this stage, we will assess our model's learning capability of understanding the context of each hand-to-mouth movement. Image-based classification of food types and calories based on images with similar techniques have previously been attempted (Kawano and Yanai, 2014; Miyazaki et al., 2011). We plan to extend this to video-based detection. Based on our dataset, we will train our model to distinguish between food and drink intake, as well as used vessel (e.g., fork, spoon, hand) and food type (lasagna, bread, yoghurt) used during the experiments. For these classifications, our models will rely on the visual differences between the gestures used in consuming different foods with different vessels. We believe that even information such as body pose during consumption gives away details about the type of food consumed. We also plan to experiment with limited caloric intake detection. Here, our models will rely on both the count of registered hand-to-mouth movements as well as the associated food types. Previous research has shown that bite count itself contains significant information about calorie consumption (Scisco et al., 2014). We expect that additional knowledge would lead to better estimates.

5 Discussion and Further Research Agenda

Dietary intake is the largest single factor contributing to disability-adjusted life expectancy in the global burden of disease (Popkin et al., 2012). In this paper, we have identified the need for UAS in the context of dietary monitoring as a facilitator for healthy nutrition. Existing systems primarily focusing on active capture methods are found to lack behavioral and interactive elements to maintain user motivation, while being perceived as burdensome. To this end, we have reviewed the status quo of active nutrition monitoring and its drawbacks, as well as current research on passive monitoring. The given framework is intended to provide a groundwork for UAS in the context of dietary monitoring. We then go on to outline a study as proof-of-concept of specific aspects of the framework. We expect the benefits of extending active food monitoring solutions by our framework to include (i) reduced amounts of under-reporting similar to previous results (Gemming et al., 2015; O'Loughlin et al., 2013) and (ii) a reduction in the perceived effort of keeping a food diary, which is often associated with guidance design features (Maedche et al., 2016; Morana et al., 2017). Going further, the informational and proactive assistance based on dietary assessment and behavioral observation should also (iii) contribute to a behavior shift towards healthier nutrition. Existing studies indicate that provision of memory aids in image format can significantly reduce under-reporting (Gemming et al., 2015), and reminders based on passive detection can lead to improved food journaling (Ye et al., 2016). However, empirical research will be required to evaluate the system, and to determine the efficacy of UAS to achieve behavior change.

In our approach, we chose computer vision-based passive monitoring as a source of information, due to its recent breakthroughs in object and action detection. Another item on our research agenda is a multimodal approach, where our models learn from both video and inertial sensor data from wrist movements. Deep learning models will initially be helpful in detecting timing (the *when*) of food intake, which will in turn support active monitoring. In the foreseeable future, monitoring of food types (*what*) and amounts (*how much*) will still require an active component, especially when considering condiments. An achievement of high detection accuracies in these tasks requires hundreds of training examples for each food. More ambitious applications of deep learning in this field will therefore require the creation of much larger databases, including a wide variety of settings, food types, eating utensils, and eating styles. This need for labelled data is one of the main challenges of deep learning today. Progress in research on unsupervised and semi-supervised learning, where only a fraction of examples need to be labelled, could reduce the dependence on the time-intensive labelling process.

Assessing eating occasions raises important concerns over end-user privacy (Purpura et al., 2011). We believe that all data should by default remain confidential, and that the user should be in control of what happens with their data. An end-user system building on our model would consist of a 360-degree camera outfitted with a processor and a trained deep learning model. While we envision all video processing to take place on the user-owned device, there will have to be an interface to communicate derived information to the UAS. Another limitation is the level of user acceptance given the need for constant video monitoring. The technology will ask the user to trust that the "black box" system uses sensor data solely for nutrition monitoring and not for any other purposes. We will explore such questions as part of our post-experiment questionnaire, to determine whether the utility of improved dietary monitoring outweighs the concerns over the presence of a video camera during eating occasions. A further limitation of surveillance video is the stationary nature of the sensor, as food consumed on the go cannot be accounted for. Hence another future direction of our research will be applying deep learning to video and accelerometer data obtained from mobile devices such as that described in Sun et al. (2014).

Acknowledgements

This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

References

Abadi, M., A. Agarwa, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, et al. (2016). "Tensorflow: Large-scale machine learning on heterogeneous distributed systems." *arXiv Preprint arXiv:1603.04467*.

Aljaroodi, H.M., M.T.P. Adam, R. Chiong, D.J. Cornforth and M. Minichiello (2017). "Empathic avatars in stroke rehabilitation: A co-designed mHealth artifact for stroke survivors." In: *Proceedings of the 2017 International Conference on Design Science Research in Information Systems (DESRIST 2017)*, pp. 73–89.

Amft, O., M. Stäger, P. Lukowicz and G. Tröster (2005). "Analysis of chewing sounds for dietary monitoring." In: *Proceedings of the 7th International Conference on Ubiquitous Computing (UbiComp 2005)*, pp. 56–72.

Ashman, A.M., C.E. Collins, L.J. Brown, K.M. Rae and M.E. Rollo (2017). "Validation of a smartphone image-based dietary assessment method for pregnant women." *Nutrients* 9 (1), 1–17.

Atkins, L. and S. Michie (2015). "Designing interventions to change eating behaviours." *Proceedings of the Nutrition Society* 74 (2), 164–170.

Australian Bureau of Statistics (2016). "Consumption Of Food Groups From The Australian Dietary Guidelines, 2011-12." *Australian Health Survey*, URL: http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4364.0.55.012main+features12011-12 (visited on 01/11/2017).

Block, G. (1982). "A review of validation of dietary assessment methods." *American Journal of Epidemiology* 115 (4), 492–505.

Burke, L.E., Y. Zheng, Q. Ma, J. Mancino, I. Loar, E. Music, M. Styn, et al. (2017). "The SMARTER pilot study: Testing feasibility of real-time feedback for dietary self-monitoring." *Preventive Medicine Reports* 6, 278–285.

Christiano, P., J. Leike, T.B. Brown, M. Martic, S. Legg and D. Amodei (2017). "Deep reinforcement learning from human preferences." In: *Advances in Neural Information Processing Systems (NIPS 2017)*, pp. 4302–4310.

Darby, A., M.W. Strum, E. Holmes and J. Gatwood (2016). "A review of nutritional tracking mobile applications for diabetes patient use." *Diabetes Technology & Therapeutics* 18 (3), 200–212.

Dong, Y., A. Hoover, J. Scisco and E. Muth (2012). "A new method for measuring meal intake in humans via automated wrist motion tracking." *Applied Psychophysiology and Biofeedback* 37 (3), 205–215.

Fogg, B.J. (2009). "Persuasive technology: Using computers to change what we think and do." In: *Proceedings of the 4th International Conference on Persuasive Technology (Persuasive'09)*, pp. 1–7.

Gemming, L., E. Rush, R. Maddison, A. Doherty, N. Gant, J. Utter and C. Ni Mhurchu (2015). "Wearable cameras can reduce dietary under-reporting: doubly labelled water validation of a camera-assisted 24 h recall." *British Journal of Nutrition* 113 (2), 284–291.

Hales, C.M., M.D. Carroll, C.D. Fryar and C.L. Ogden (2017). *Prevalence of Obesity Among Adults and Youth: United States, 2015-2016*. NCHS Data Brief No. 288. Hyattsville, MD: National Center for Health Statistics.

Hochreiter, S. and J.J. Schmidhuber (1997). "Long short-term memory." *Neural Computation* 9 (8), 1735–1780.

Kawano, Y. and K. Yanai (2014). "Food image recognition with deep convolutional features." In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2014)*, pp. 589–593.

Krebs, P. and D.T. Duncan (2015). "Health app use among US mobile phone owners: A national survey." *JMIR mHealth and uHealth* 3 (4), 1–12.

Krizhevsky, A., I. Sulskever and G.E. Hinton (2012). "ImageNet classification with deep convolutional neural networks." In: *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pp. 1097–1105.

LeCun, Y., Y. Bengio and G. Hinton (2015). "Deep learning." *Nature* 521, 436–444.

Lee, J.-E., S. Song, J. Ahn, Y. Kim and J. Lee (2017). "Use of a mobile application for self-monitoring dietary intake: Feasibility test and an intervention study." *Nutrients* 9 (7), 1–12.

Lichtman, S.W., K. Pisarska, E.R. Berman, M. Pestone, H. Dowling, E. Offenbacher, H. Weisel, et al. (1992). "Discrepancy between self-reported and actual caloric intake and exercise in obese subjects." *The New England Journal of Medicine* 327 (27), 1893–1898.

Locke, E.A. and G.P. Latham (2002). "Building a practically useful theory of goal setting and task motivation: A 35-year odyssey." *American Psychologist* 57 (9), 705–717.

Maedche, A., S. Morana, S. Schacht, D. Werth and J. Krumeich (2016). "Advanced user assistance systems." *Business & Information Systems Engineering* 58 (5), 367–370.

Michie, S., M.M. van Stralen and R. West (2011). "The behaviour change wheel: A new method for characterising and designing behaviour change interventions." *Implementation Science* 6, 1–11.

Miyazaki, T., G.C. De Silva and K. Aizawa (2011). "Image-based calorie content estimation for dietary assessment." In: *Proceedings of the 2011 IEEE International Symposium on Multimedia (ISM 2011)*, pp. 363–368.

Morana, S., S. Schacht, A. Scherp and A. Maedche (2017). "A review of the nature and effects of guidance design features." *Decision Support Systems* 97, 31–42.

O'Loughlin, G., S.J. Cullen, A. McGoldrick, S. O'Connor, R. Blain, S. O'Malley and G.D. Warrington (2013). "Using a wearable camera to increase the accuracy of dietary analysis." *American Journal of Preventive Medicine* 44 (3), 297–301.

Oinas-Kukkonen, H. and M. Harjumaa (2009). "Persuasive Systems Design: Key Issues, Process Model, and System Features." *Communications of the Association for Information Systems* 24 (3), 485–500.

Popkin, B.M., L.S. Adair and S.W. Ng (2012). "Global nutrition transition and the pandemic of obesity in developing countries." *Nutrition Reviews* 70 (1), 3–21.

Purpura, S., V. Schwanda, K. Williams, W. Stubler and P. Sengers (2011). "Fit4Life: The design of a persuasive technology promoting healthy behavior and ideal weight." In: *Proceedings of the 2011 SIGCHI Conference on Human Factors in Computing Systems*, pp. 423–432.

Sasaki, S., A. Katagiri, T. Tsuji, T. Shimoda and K. Amano (2003). "Self-reported rate of eating correlates with body mass index in 18-y-old Japanese women." *International Journal of Obesity* 27, 1405–1410.

Scisco, J.L., E.R. Muth and A.W. Hoover (2014). "Examining the utility of a bite-count-based measure of eating activity in free-living human beings." *Journal of the Academy of Nutrition and Dietetics* 114 (3), 464–469.

Solbrig, L., R. Jones, D. Kavanagh, J. May, T. Parkin and J. Andrade (2017). "People trying to lose weight dislike calorie counting apps and want motivational support to help them achieve their goals." *Internet Interventions* 7, 23–31.

Sun, M., L.E. Burke, Z.-H. Mao, Y. Chen, H.-C. Chen, Y. Bai, Y. Li, et al. (2014). "EButton: A wearable computer for health monitoring and personal assistance." In: *Proceedings of the 2014 51st Design Automation Conference (DAC)*, pp. 1–6.

Turner-McGrievy, G.M., M.W. Beets, J.B. Moore, A.T. Kaczynski, D.J. Barr-Anderson and D.F. Tate (2013). "Comparison of traditional versus mobile app self-monitoring of physical activity and dietary intake among overweight adults participating in an mHealth weight loss program." *Journal of the American Medical Informatics Association* 20 (3), 513–518.

Vu, T., F. Lin, N. Alshurafa and W. Xu (2017). "Wearable food intake monitoring technologies: A comprehensive review." *Computers* 6 (1), 1–28.

Westerterp, K.R. and A.H. Goris (2002). "Validity of the assessment of dietary intake: problems of misreporting." *Current Opinion in Clinical Nutrition and Metabolic Care* 5 (5), 489–493.

WHO (2017). "Noncommunicable Diseases." *World Health Organization Media Centre*, URL: http://www.who.int/mediacentre/factsheets/fs355/en/ (visited on 01/11/2017).

Ye, X., G. Chen, Y. Gao, H. Wang and Y. Cao (2016). "Assisting food journaling with automatic eating detection." In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA'16)*, pp. 3255–3262.

Zhu, F., M. Bosch, I. Woo, S. Kim, C.J. Boushey, D.S. Ebert and E.J. Delp (2010). "The use of mobile devices in aiding dietary assessment and evaluation." *IEEE Journal of Selected Topics in Signal Processing* 4 (4), 756–766.