Published in final edited form as:
J Meas Phys Behav. 2018 September ; 1(3): 143–154. doi:10.1123/jmpb.2018-0021.

Unique Views on Obesity-Related Behaviors and Environments: Research Using Still and Video Images

Jordan A. Carlson,
Children's Mercy Kansas City

J. Aaron Hipp,
North Carolina State University

Jacqueline Kerr,
University of California San Diego

Todd S. Horowitz, and
National Cancer Institute

David Berrigan
National Cancer Institute

Abstract

Objectives: To document challenges to and benefits from research involving the use of images by capturing examples of such research to assess physical activity– or nutrition-related behaviors and/or environments.

Methods: Researchers (i.e., key informants) using image capture in their research were identified through knowledge and networks of the authors of this paper and through literature search. Twenty-nine key informants completed a survey covering the type of research, source of images, and challenges and benefits experienced, developed specifically for this study.

Results: Most respondents used still images in their research, with only 26.7% using video. Image sources were categorized as participant generated (n = 13; e.g., participants using smartphones for dietary assessment), researcher generated (n = 10; e.g., wearable cameras with automatic image capture), or curated from third parties (n = 7; e.g., Google Street View). Two of the major challenges that emerged included the need for automated processing of large datasets (58.8%) and participant recruitment/compliance (41.2%). Benefit-related themes included greater perspectives on obesity with increased data coverage (34.6%) and improved accuracy of behavior and environment assessment (34.6%).

Conclusions: Technological advances will support the increased use of images in the assessment of physical activity, nutrition behaviors, and environments. To advance this area of

Carlson (jacarlson@cmh.edu) is corresponding author.
Carlson is with the Center for Children's Healthy Lifestyles and Nutrition, Children's Mercy Kansas City, Kansas City, MO. Hipp is with the Dept. of Parks, Recreation, and Tourism Management, College of Natural Resources, and a Fellow at the Center for Geospatial Analytics, North Carolina State University, Raleigh, NC. Kerr is with the Dept. of Family Medicine and Public Health, Moores Cancer Center, University of California San Diego, La Jolla, CA. Horowitz and Berrigan are with the Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD.
Carlson and Hipp are co-first authors.

research, more effective collaborations are needed between health and computer scientists. In particular development of automated data extraction methods for diverse aspects of behavior, environment, and food characteristics are needed. Additionally, progress in standards for addressing ethical issues related to image capture for research purposes is critical.

Keywords

physical activity; computer vision; engineering; machine learning; measurement; nutrition

Systematic observation of health behaviors in context is the gold standard in public health research on built environments and obesity-related behaviors (Evenson, Jones, Holliday, Cohen, & McKenzie, 2016; Glanz, Sallis, Saelens, & Frank, 2007; Joseph & Maddock, 2016). Traditional standards place researchers or trained community members physically in the environment with a set protocol (McKenzie, Cohen, Sehgal, Williamson, & Golinelli, 2006). Such research is resource intensive, especially in personnel, and introduces limitations including frequency and duration of time during which an environment and behavior may be observed. Auditing of environments supporting physical activity and healthy eating is moving from pen and paper observational audits to digital audits and online evaluation via Google Street View (GSV) (Bader, Mooney, Bennett, & Rundle, 2017; Bader et al., 2015; Eyler et al., 2015). Additional emerging technologies, including the improvements in the quantity, quality, and diversity of camera and video devices, have greatly expanded the potential sources and analytic techniques for systematic observation (Graham & Hipp, 2014; Loveday, Sherar, Sanders, Sanderson, & Esliger, 2015; Park & Ewing, 2017). This expansion in methods coincides with recent recommendations from the Centers for Disease Control and Prevention and the American College of Sports Medicine for an overarching physical activity research and surveillance strategy that prioritizes alternative sources of data, including investigating device-based assessment technologies (Fulton et al., 2016).

Images have been captured for evaluative health purposes since at least the advent of the X-Ray in 1895 (Spiegel, 1995) and the annotation of human behavior in public spaces via video since at least William ‘Holly’ Whyte’s pioneering work in New York City plazas during the 1970s (Whyte, 1980). Fast forward to 2018 where built environment and physical activity/healthy eating research utilizing image capture from Unmanned Aerial Vehicles (UAV or drones), smartphone applications (apps), webcams, and other modalities is emerging, advancing public health research and methodology in new directions. Park and Ewing (2017), for example, detail the benefits and challenges of using UAV in completing the System for Observing Play and Recreation in Communities (SOPARC) in parks. The benefits include the ability to systematically observe larger spaces and the ability to save, review, and validate video to ensure correct numbers and intensity of physical activity. Further, archived records allow review of behaviors when new discoveries emerge, for example including observations of sedentary behavior as well as physical activity. In the field of nutrition, Cowburn and colleagues (Cowburn et al., 2015) have recently used wearable cameras to track teenagers as they commute to and from school. With the cameras capturing an image every 15 seconds, researchers were able to determine environmental exposures to food and food marketing and corresponding intake decisions. The built environment is also being measured with captured images and video, often through GSV. Bader et al. (2017), used GSV images to compare disorder across neighborhoods in large US cities while also providing support for the use of GSV in conducting walkability audits, studying gentrification, and observing the effects of natural disasters. Table 1 provides additional examples of different image types for the collection of physical activity, diet, and environments.

Similar to the advances in spatial energetics that have been possible due to new Global Positioning Systems (GPS) and Geographic Information Systems (GIS) technologies (James et al., 2016), we believe that image- and video-based data collection could contribute important findings and methodological advances to the field of obesity prevention. Given the variety of image/video-capture devices (e.g., person-worn/SenseCam, car-top/GSV, street intersection/traffic cameras, UAV) and differences in researcher perspectives using images (e.g., public health, computer science, anthropology), this paper aims to provide a guided overview and discussion of current efforts. We sought to gain practical information beyond the published literature by contacting 51 researchers and asking them a series of questions focused on the use of images and video to assess behaviors and environments related to obesity. Similar to the GPS example above, practical information at the beginning of a new methodological era is helpful to researchers (Kerr, Duncan, & Schipperjin, 2011). Behaviors solicited included physical activity, sedentary behavior, transportation, food purchasing, and food consumption. Environments included home, work, and neighborhood. Still images and video data could be captured by person-worn cameras, mobile devices, or stationary cameras. We aimed to uncover and highlight methodological and practical challenges of image data collection and provide examples of successes to inspire further efforts at image based discovery. Not only is image data collection challenging due to its technical nature, but images are also considered Protected Health Information, so are subject to additional privacy concerns. These and other challenges were probed.

Methods

Participants

Key informants were 29 researchers (30 response forms, one researcher completed two forms capturing two distinct research topics) out of 51 contacted (56.9% response rate). The 51 eligible researchers were those believed to be using image capture in their research, based on the authors' knowledge and published materials. Google Scholar and PubMed were also used to identify such researchers, using the search terms "street view," "photovoice," "video direct observation," and "images and diet assessment." Key informants were also asked to suggest colleagues working with image capture. The purpose of the data collection was not to be exhaustive but rather to document examples of how image capture was being used in current research, and the challenges and benefits to using image capture in research. The data collection form was administered through REDCap and sent via email in early 2017, with one follow up email and phone call prompt, as necessary. This study received ethical approval from the human subjects' protections committee at Children's Mercy Kansas City. De-identified datasets for the current study are available from the corresponding author on reasonable request.

The evaluation did not include clinical image analysis such as computerized tomography scans or patient-provider interactions. It did not include mobile health (mhealth) interventions where images are delivered, but not collected from participants or in their environments. Use of images from media such as advertising was excluded, but image assessment of advertisement exposure was included (e.g., sugar sweetened beverage advertisements along usual walking route as captured by person-worn camera or GSV). Use of remote sensing, e.g., webcams or use of commercial, ground-based image databases such as GSV were included, but satellite imagery was excluded (e.g., MODIS for land cover analysis). Finally, photovoice (Davis, Goldmon, & Coker-Appiah, 2011; Wang & Burris, 1997) was included in our initial scoping of relevant literature and researchers.

Twenty of the key informants were from the US, with nine from other countries including Australia, Belgium, Canada, Chile, Columbia, England, and New Zealand. The respondents were trained in diverse disciplines including medicine (2), architecture and planning (2), computing (1), geography (1), clinical psychology (3), public health (6), nutrition (2), kinesiology (3) and other topics. Most (*n* = 19) of the respondents were in health oriented University Departments such as Public Health, Epidemiology, Gerontology, Nutrition and Kinesiology. The remainder respondents worked at hospitals (1), government agencies (2), or private research centers (2).

Data Collection Tool

The data collection tool included multiple questions about the type of research, the challenges faced, and the benefits experienced from using image capture. One question asked the name/description of the device for collecting image/video data, and was used to categorize the types of image collection covered as (1) participant generated: images taken by research participants; (2) researcher generated: images captured under the direction of the researcher or automatically by the camera; and (3) curated from third party: images publicly available and leveraged by the investigator for research purposes.

Key informants reported on the obesity-related constructs that were included in their research, categorized as physical activity (including sedentary behavior) and/or nutrition, and whether physical activity/nutrition behaviors, environments, or both were investigated. Other questions included the image frame (still/single frame vs. video/many frames), current research classification (development, validation, correlational [e.g., investigating associations with health/behavior/environments], intervention, and evaluation [e.g., of naturalistic interventions]), settings covered (home, work, school, parks/recreation, neighborhood, other buildings, social interaction, advertising, and social media), populations covered (children, teenagers, adults, older adults, patients, and special populations), and research stage (pilot, funded, in progress, completed, and published). The aforementioned response options were not mutually exclusive, so the respondent could check multiple options.

Key informants were asked to check each of the challenges they faced in their research involving image capture from a list of seven challenges (receptivity/recruitment, data collection, data processing, data quality, ethical issues, participant burden, and scalability). Lastly, two open-ended questions asked investigators to describe (1) the top challenges they have faced, and (2) the top benefits they have experienced in using image capture in their research.

Analysis

Descriptive statistics were used to present frequencies of responses overall, by device/platform, and by image collection category. Inductive thematic analysis was used for the two qualitative questions to identify emerging themes and frequencies of occurrence. The thematic analysis was performed by one author and confirmed by the other authors of this paper.

Results

Information on the image frame, current research classifications, settings covered, populations covered, and research stage for the included sample are presented in the Appendix. Most (93.3%) key informants who responded were using still images, with 26.7% using video (note some percentages will exceed 100% due to multiple methods within studies). The research classifications were fairly balanced across development (46.7%), validation (56.7%), correlational (46.7%), intervention (30.0%), and evaluation (33.3%). The settings covered were also balanced across home (36.7%), work (26.7%), school (33.3%), parks/recreation (53.3%), neighborhood (70.0%), other buildings (23.3%), and social interaction (30.0%). A good representation of research involving children (46.7%), teenagers (46.7%), adults (73.3%), and older adults (43.3%) was captured. Half (50.0%) of the research included had been published.

Image and Video Capture

The participant-generated image category (*n* = 13) consisted of research using smartphones to capture nutrition-related images or photovoice-type methods. Photovoice-type methods included community members capturing images of built environment attributes to support advocacy efforts for improvements. The researcher generated category (n = 10) consisted of research using wearable cameras (e.g., SenseCam) or stationary cameras. The curated from third party category (*n* = 7) consisted primarily of research using GSV and/or Google Earth but also included one respondent using images from social media and another using images from publicly available webcams (Table 1). Curated images were primarily focused on parks/recreation (57.1%) and neighborhoods (85.7%).

Over 80% of the image capture research involved physical activity behavior or environments, while just under 40% involved nutrition behavior or environments (Figure 1). Use of GSV was specific to physical activity environment research, use of stationary cameras was specific to physical activity behavior research with one exception (out of 5) which also assessed the physical activity environment, and use of smartphone images was specific to nutrition research (Figure 2). The frequencies are presented simply to describe the sample and range of research perspectives, they are not meant to represent the prevalence of ongoing research.

Challenges

Data processing was the most frequently endorsed challenge from the provided list of challenges to using image capture in research, particularly in researcher generated studies (endorsed by 63.3% of all respondents and 100% of those using researcher generated image capture; Table 2). Data collection was the most frequently endorsed challenge for those using participant generated image capture (53.8%) and images curated from a third party (85.7%).

Seven themes emerged from the open item asking key informants to report the greatest challenges they faced in their image capture research (Table 3).

The most frequently occurring theme was the need for automated processing (mentioned by 58.8% of respondents), which was particularly relevant to those using researcher generated image capture (85.7%) and those using images curated from a third party (75.0%). Quality of video/images was endorsed less by those in the researcher generated category (14.3%) as compared to the other image type categories (50.0%), whereas scalability was endorsed the most by those in the researcher generated category (42.9% vs. 0–16.7%). Selected quotes regarding challenges to using image capture in research are presented below (with corresponding theme):

- "The largest burden for our SenseCam studies has been the time and energy needed to annotate the images—it takes about 3 hours to annotate all of our position, activity, and environment labels for each participant day collected." (need for automated processing)
- "Tracking pedestrians using computer vision techniques in urban environments is not trivial, too much visual noise." (need for automated processing)
- "Sometimes [Street View] imagery is unavailable or of poor quality for the desired date … [and] it is challenging to assess subjective features using imagery (e.g., aesthetics, maintenance, sidewalk quality, indicators of physical disorder and the social environment)." (quality of video/images and device/platform challenges)
- "I have been collaborating, but this is not always easy. It takes time. There has to be a win and an improvement in science for both of us." (successful collaborations)
- "Sites don't want a camera recording their activities." (participant challenges)
- "Our university IRB has strict concerns with the 'bystanders' captured in the images we collect. There have been several requirements we have been able to accommodate (i.e., encrypting the cameras so participants cannot access the images at home)." (ethics/IRB)

Benefits

Ten themes emerged from the item asking key informants to report the largest benefits they experienced by using image capture in their research (Table 4).

The most frequently occurring themes for those using participant generated image capture were that the image capture supported participant engagement/intervention (endorsed by72.7% of respondents) and generated qualitative data to complement quantitative data (54.5%). The most frequently occurring theme for those using researcher generated image capture was that image capture improved accuracy of measurement (87.5%), and the most frequently occurring themes for those using images curated from a third party were that using image capture reduced researcher burden (57.1%) and provided more data (71.4%). Selected quotes regarding benefits to using image capture in research are presented below (with corresponding theme):

- "It [can] take away the subjectivity and burden of human observers." (automate, reduce researcher burden/make feasible, and improve accuracy)
- "The additional information capture in the before and after images of an eating occasion has provided new context about eating not known before, e.g., detailed time, location. In turn, this has generated new research questions and new ways to evaluate diet, e.g., sustainability." (provides more data)
- "It allows us to be in more places at once. And it allows us to collect data 24 hrs. a day." (provides more data)
- "The resulting images were very telling and increased awareness in the community of issues to be addressed." (participant engagement/intervention tool and qualitative data to complement quantitative data)
- "It has provided a 'ground truth' data set … [that] has enabled us to develop algorithms that can be applied to epidemiological cohorts of free living populations." (validation/algorithm training)
- "Real-time analysis can be used to trigger actuations in urban infrastructures." (real-time or rapid feedback)

Discussion

A primary aim of the present study was to document challenges, opportunities, and successes in the use of images and video for the collection of critical data connected to physical activity and diet-related behavior. Direct observations of physical activity and diet are the gold standards for measurement of these behaviors and offer the added advantage of facilitating simultaneous measures of built, natural, and social environments in which these behaviors occur. The results of this study captured the enthusiasm felt by researchers concerning the potential of data collection via images, but also highlighted several barriers, most notably, the need for advances in automated processing of images, for pathways to collaboration with computer scientists, and for standards for dealing with privacy concerns.

Image-based context measures can help validate other sensor measures in real world settings outside of the laboratory (Kerr et al., 2013), provide large scale behavioral measurement in place (Hipp et al., 2017), and provide additional contextual information not previously available to generate new research questions (Colabianchi, 2015; Hipp et al., 2017). For example, as more interventions measure environmental change and its subsequent impact on behavior, and as Just in Time Adaptive Interventions through mobile devices attempt to change individual behaviors in place, valid measures of behaviors in context are key (Nahum-Shani, Hekler, & Spruijt-Metz, 2015; Riley et al., 2011). Some settings and behaviors, e.g., cycling, are easier to monitor than others due to existing camera infrastructure and the uniqueness of the observable behavior (Hipp, Adlakha, Eyler, Chang, & Pless, 2013). Other measures such as calories may not be identifiable by images alone, either because the content is masked or because healthy and unhealthy versions look identical (Boushey, Spoden, Zhu, Delp, & Kerr, 2016). In such situations, images can still be useful in conjunction with experts or used to aid or prompt recall in participants.

Automated Processing

Researchers repeatedly mentioned the need for automated image processing. Availability of software tools that accelerate extraction of behavioral, contextual, and compositional data could increase the use of images for data collection. Currently, very few tools that can process the scale of population level data employed in public health are available. When available, tools are either developed ad hoc or public health and computer science and engineering researchers have been unable to move forward with inter- or trans-disciplinary problem solving. Efforts have been made to develop software to characterize food intake based on intermittent photos from person-worn cameras or from photos taken by study subjects of meals (Boushey et al., 2017; Boushey et al., 2016; Martin et al., 2009). While these and other related tools have been developed to apply to diverse study designs and measurement challenges, results have been mixed as to whether they reduce measurement costs or increase validity. To our knowledge, there are no commercially available image analysis tools for dietary assessment. Computer aid has helped accelerate data extraction from sequences of images, but it has not eliminated the need for human judgment concerning behavioral content. Extraction of meaningful dietary data from images seems even more difficult. Images of food may be useful in cataloging instances of eating, but again human judgement is required to determine what on the plate or in the scene is actually being eaten and what volume.

Behavioral science has traditionally relied on labor-intensive coding methods where a small corps of research assistants is intensively trained on the behaviors or other features of interest. This strategy works well when the image datasets are relatively small, such as a few dozen videos of interviews. As the magnitude of the image datasets increase, different strategies become necessary. The field of computational vision is a source for potential solutions to this problem (Pless & Souvenir, 2009).

There are two basic strategies for efficiently coding large image datasets: crowdsourcing and computational strategies. Crowdsourcing techniques involve the researcher putting the image sets online and asking people to code them. This strategy replaces a small number of highly-trained and (relatively) well-paid study staff with a large number (the “crowd”) of untrained volunteers or pieceworkers on, for example, Amazon Mechanical Turk or CrowdFlower. While crowdsourcing has privacy and ethical implications, using publicly available images or blurring faces can address some of these concerns. These matters are discussed in more detail in the Participants, Ethics, and Coverage section below. At least one of our respondents indicated that they were employing a crowdsourcing strategy to annotate images (Hipp et al., 2017). Computational techniques, in contrast, tend to minimize the need for human intervention. While in some cases, there may be algorithms already well-suited to the scientific question the researcher is interested in, often existing algorithms will have to learn how to classify images to the researchers' specifications (Moghimi, Kerr, Johnson, Godbole, & Belongie, 2015). These strategies should be considered complementary, rather than mutually exclusive. For many research projects, the best solution may be to combine crowdsourced image annotations with computational methods, using a small team of coders to vet the outputs. Here we discuss a sampling of techniques from outside of the public health domain that may provide useful inspiration for researchers planning a project with an intensive image collection component.

Existing Tools in Other Domains.—Defense-related applications and commercial security teams are ahead of the public health research community in both categorizing human behavior and event recognition. The DARPA Mind's Eye Program was designed to support development of tools to recognize specific behaviors from video recording in real or near real time (Barrett, Xu, Yu, & Siskind, 2016). For example, the military has an interest in automatically detecting if someone is trying to climb a fence. However, these advances are not yet found in practical tools for health and environment researchers. Efforts are needed to connect the public health user community and developers and engineers engaged in more sophisticated image recognition research, as discussed in more detail in the Collaboration section below.

Examples of tools from computer imaging and informatics research that have yet to be adapted for public health research include the LabelMe tool (Torralba, Russell, & Yuen, 2010) and the Body Talk approach (Streuber et al., 2016). LabelMe is an online tool that allows users to segment and label objects in an image. There are tens of thousands of images in the dataset, which is constantly growing. Users can label as few or as many images as they like, and they do not have to label every object in every image; the power of the tool comes from aggregating responses across thousands of users. The object annotations have been used for a wide range of applications, such as testing human observers' statistical intuitions about the frequencies of objects in the world (Greene, 2016). The LabelMe approach has been extended to video data as well (Jenny, Russell, Ce, & Torralba, 2009). A researcher could use LabelMe, or a similar tool, to identify, for example, fastfood advertising from random pictures taken along students' walks to school, without ever having to explicitly instruct coders on what to look for. Similarly, Streuber et al.'s (2016) Body Talk is a combination of crowdsourcing with computer graphics and principle component analysis able to recover 3D body shape from verbal descriptions of human bodies. A similar approach could be used to take video from a fixed location, such as a park, crowdsource descriptions of the people who walk by the camera, and recover the distribution of body size of people who frequent this park.

Machine Learning or Human Computation.—Machine learning techniques might be useful for automated image classification. For example, random forest algorithms can be used to classify images of food, perhaps increasing the efficiency of processing respondent collected food photos or summaries of print or web food image environments (Bossard, Guillaumin, & Van Gool, 2014). The most prominent among current approaches are the various flavors of deep learning (LeCun, Bengio, & Hinton, 2015) algorithms. Deep learning grew out of the neural network or connectionist research project that came to prominence in the 1980s (Medler, 1998). A good example is the eight-layer network developed by Krizhevsky, Sutskever, and Hinton (2017). The network was trained on 1.2 million images classified (by humans) into 1,000 different categories, and tested on a separate set of 150,000 images. It managed to pick the correct image label for the test set with 62.5% accuracy, and 83% of the time the correct label was in the network's top five possibilities. This example nicely illustrates the enormous potential of deep neural networks, as well as the challenges for health behavior researchers. Deep neural networks require huge datasets for training, such as the 15 million-item ImageNet (Deng et al., 2009). A lot of computing power is required; the Krizhevsky et al. network required two NVIDIA GTX 580 3GB graphics processing units (touted as "The world's fastest GPUs at the time of their release) running for six days.

There are also computational approaches intended to capture aspects of human visual cognition. For example, if a researcher were interested in the proportion of images from a participant's image stream that were taken outdoors versus indoors, she might turn to the spatial envelope techniques developed by Oliva and Torralba (2006). The spatial envelope, or gist descriptors, can categorize a scene based on global image features. In other words, there is no need to segment the image and identify the individual objects. Instead, the spatial envelope is a description of the scene based on a statistical abstract of visual features similar to those known to be analyzed in the early stages of the human visual system. These descriptions can discriminate between scenes that are open or closed, more natural or more artificial, and so forth.

Event Segmentation and Action Recognition.—The approaches described above apply mostly to still frames. Techniques for analysis of video streams are less well-developed. Recent work, based on psychological principles of event perception (Zacks & Swallow, 2007), has demonstrated that the data stream from a modern smartphone, with multiple parallel sensor records (e.g., images, audio, GPS locations), can be accurately segmented into events based on analysis of accelerometer data (Zhuang, Belkin, & Dennis, 2013). However, such techniques do not provide a description of what these events might be (A fast-food meal? A jog in the park?). Here, older techniques for automatic storyboarding (Macer, Thomas, Chalabi, & Meech, 1996) could be applied to generate a summary frame for each event. This would substantially reduce the burden on the human coder; instead of watching hours of monotonous video, she would only have to classify a much smaller set of still images.

Recognizing human actions is perhaps further along than event segmentation (Rui & Anandan, 2000). It has been known for a long time that humans can correctly perceive actions when given only a dozen or so moving dots to represent the human body (Johansson, 1973) (e.g., "point-light walkers" (Thornton, Rensink, & Shiffrar, 2002)). A large field of computerized action recognition has built on this insight. There are a variety of methods for segmenting and labeling human actions, such as might be obtained from surveillance-type cameras (Thornton et al., 2002). These techniques produce reasonable accuracy in extracting and labeling basic full-body human actions (e.g., walking, running, doing jumping jacks (Zhang, Hu, Chan, & Chia, 2008)). However, these techniques are generally limited to recognizing the full-body actions of a single human being; identifying the actions of multiple people in a single video stream is challenging, and identifying interactions between people even more so. As with algorithms to characterize food images, perhaps semi-automated recognition of human actions could accelerate processing of image files collected to describe physical activity in outdoor settings such as parks and playgrounds (Carlson et al., 2017). Several of our respondents (*n* =5) were capturing images from cameras worn by participants. Methods for analyzing such egocentric video content are being developed. For example, Vaca-Castano et al. (2017) combine insights from human scene recognition with machine learning to recognize objects in egocentric videos of people carrying out activities of everyday living (Ramanan, 2012). These techniques capitalize on the visual context, such as the fact that an object is more likely to be a microwave if it is in a kitchen scene and not a bedroom.

It is not yet possible to purchase a software package off the shelf that would allow automatic coding of images for research purposes. However, depending on the research question, there may be some combination of computational and crowdsourcing techniques that could substantially ease the burden on coders, and facilitate the development of this research field. Collaborations between public health researchers and computational vision specialists should be encouraged. Such collaboration might help guide computer science advances that would be immediately useful for applied research on public health. Martin et al. pointed out that “Two central problems in vision are image segmentation and recognition. Both problems are hard, and we do not yet have any general purpose solution approaching human level competence for either one” (Martin, Fowlkes, Tal, & Malik, 2001). It is not clear that either problem has been solved to date, but there may be partial solutions of use to health researchers.

Collaboration

While health experts possess knowledge of the health implications and real-world application potential for a given technology, they rarely have the expertise needed for developing or customizing technologies (e.g., computer vision algorithms). Creating successful collaborations, often between health experts and computer scientists or engineers, emerged as a critical factor for success in image-based research projects, and several challenges to these collaborations were noted. One challenge is that terminology differs substantially between fields, so each researcher needs to learn the terminology of their collaborator. Another challenge is that there is currently little integration between the fields of health and computer science/engineering (e.g., different academic departments, separate conferences, distinct journal outlets), but this appears to be improving. This lack of integration is apparent in the sample of the present study, which consisted primarily of health researchers with only a handful of computer scientists and engineers represented. Considerations in such collaborations include addressing the different motivators and desired outcomes of each researcher, for example framing the research question/problem in a way that excites all team members (e.g., addressing health implications as well as innovation/

advancement in computer vision) and publishing findings in both health and computer science/engineering outlets. The latter can be challenging because journal editors and reviewers look for quite different qualities when gauging merit of a research project (e.g., health implications and application vs. technological innovation and rigor). While technological research is being incorporated into health conferences at a fairly rapid rate (e.g., in the Society of Behavioral Medicine), such conferences still tend to draw very few computer scientists and engineers. More efforts should be made to merge conferences and societies drawing these distinct disciplines. One key lesson from our own work and from these interviews is that collaboration across diverse discipline takes a significant commitment of time and energy (Hall et al., 2012).

Participants, Ethics, and Coverage

Lastly, respondents raised questions concerning ethics, coverage, and participant engagement. We know of one IRB that did not permit automated data processing so that participants themselves had to annotate the images. Most of the concerns arise from the ability of cameras to capture images of participants not consented into the research process. However, much of this concern is misplaced as images of members of the public without any identifying information does not elevate them to third party members at risk. Many efforts are made to protect participant images, including blurring of images when stored on the device or masking of faces or identifiable features post processing, guiding participants on when image capture is appropriate (for example not in changing rooms), allowing participants to delete images, and providing participants with information to help them explain the research to others. Studies have shown that participants understand the consenting process for image capture studies and that some IRB concerns are not shared by participants (Nebeker et al., 2016). As more research demonstrates the benefits of image data collection, it may be easier to present more persuasive arguments to IRBs concerning the balance between benefit and risk. Efforts have been made by researchers to provide examples and arguments to support successful image-based research applications and appropriate participant protections. One such effort is the Connected and Open Research Ethics (CORE) platform (https://thecore.ucsd.edu/). Another useful effort is the ethical framework for supporting research using wearable cameras that was developed by health behavior researchers (Kelly, Marshall, et al., 2013).

Other ethical concerns include image capture of dangerous or illegal behavior. As of 2017, the National Institutes of Health automatically issues a Certificate of Confidentiality for all grants that involve identifiable and sensitive information. For projects with other funding sources in the U.S., researchers can apply for a Certificate of Confidentiality (National Institutes of Health, 2017). A Certificate of Confidentiality protects participants against legal actions (e.g., based on the image content), restricting the disclosure of identifiable, sensitive information outside of the study team unless required by other Federal, State, or local laws, such as for reporting of communicable diseases; unless the subject consents; or for the purposes of scientific research that is compliant with human subjects regulations. However, researchers also need to comply with requirements of medical professionals, for example to report child or elder abuse (e.g., (U.S. Dept. of Health and Human Services, 2016)).

Images are considered Protected Health Information, so additional security procedures are required. With large image sets this can lead to additional data storage costs. In some cases researchers may only be able to share annotations of images (not the images themselves). However, many studies would benefit from sharing images and having other interested research groups aid in the annotation process or algorithm development, a common practice in computer science. Further, in computer vision the feature creation process is often where new developments are made that advance algorithm performance. Researchers would need access to the original images to create such innovative features. It has also been our experience that standard IRB training does not prepare transdisciplinary groups for protecting images and additional steps are required to highlight how the data can be stored and accessed.

Conclusions

Key informants, largely from the public health sector, shared experiences and challenges related to data collection from still and moving images for obesity research. Successful projects tended to include researchers from multiple disciplines. Keys to successful collaborations include joint goals, creation of a common language, and respect for each other's expertise. While engineers can provide technical support and solutions, health researchers provide insight into human research and the design of studies to validate or test the boundaries of measures that are relevant to public health. Some challenges of this research include the need for human annotations, at least at early stages of the research or more complex recognition tasks. Many researchers are using online crowdsourcing tools for their scalability and immediacy. However, more local community members may also provide important perspective and image analysis might further engage communities in environmental advocacy. Given progress in collaboration across disciplines, advancing automation of data extraction from images and clear guidelines concerning ethical issues are the major challenges to fulfilling the potential of data collection concerning obesity and its behavioral and environmental determinants from images. The use of images and image/video analysis in public health research has multiple benefits for advancing the science of physical activity and nutrition. The many successes reported in this paper illustrate the promise that a greater focus on development and application of image-based assessment methods will yield tangible increases in our understanding of health behaviors in context.

Acknowledgments

We thank the many researchers who shared their work and viewpoints with us.

Appendix

Table A1

Descriptive information on the types of research covered in sample

		***N* (%) by Image Collection Category**		
	Total *N* (%) (*N* = 30)	**Participant Generated (*n* =13)**	**Researcher Generated (*n* =10)**	**Curated from Third Party (*n* = 7)**
Image frame				
Still/single frame	28 (93.3%)	13 (100%)	8 (80.0%)	7 (100%)
Video/many frames	8 (26.7%)	3 (23.1%)	5 (50.0%)	0
Current research classification				
Development	14 (46.7%)	5 (38.5%)	6 (60.0%)	3 (42.9%)
Validation	17 (56.7%)	6 (46.2%)	7 (70.0%)	4 (57.1%)
Correlational	14 (46.7%)	7 (53.8%)	3 (30.0%)	4 (57.1%)
Intervention	9 (30.0%)	8 (61.5%)	0	1 (14.3%)
Evaluation	10 (33.3%)	6 (46.2%)	2 (20.0%)	2 (28.6%)
Settings covered				
Home	11 (36.7%)	7 (53.8%)	3 (30.0%)	1 (14.3%)
Work	8 (26.7%)	4 (30.8%)	4 (40.0%)	0
School	10 (33.3%)	5 (38.5%)	5 (50.0%)	0
Parks/recreation	16 (53.3%)	6 (46.2%)	6 (60.0%)	4 (57.1%)
Neighborhood	21 (70.0%)	9 (69.2%)	6 (60.0%)	6 (85.7%)
Other buildings	7 (23.3%)	3 (23.1%)	3 (30.0%)	1 (14.3%)
Social interaction	9 (30.0%)	6 (46.2%)	3 (30.0%)	0
Advertising	1 (3.3%)	0 (0%)	1 (10.0%)	0
Social media	3 (10.0%)	1 (7.7%)	1 (10.0%)	1 (14.3%)
Populations covered				
Children	14 (46.7%)	6 (46.2%)	4 (40.0%)	4 (57.1%)
Teenagers	14 (46.7%)	6 (46.2%)	3 (30.0%)	5 (71.4%)
Adults	22 (73.3%)	8 (61.5%)	9 (90.0%)	5 (71.4%)
Older adults	13 (43.3%)	6 (46.2%)	4 (40.0%)	3 (42.9%)
Patients	3 (10.0%)	1 (7.7%)	2 (20.0%)	0
Special populations	7 (23.3%)	6 (46.2%)	1 (10.0%)	0
Research stage				
Pilot	5 (16.7%)	4 (30.8%)	0	1 (14.3%)
Funded	15 (50.0%)	9 (69.2%)	5 (50.0%)	1 (14.3%)
In progress	16 (53.3%)	7 (53.8%)	4 (40.0%)	5 (71.4%)
Completed	18 (60.0%)	8 (61.5%)	5 (50.0%)	5 (71.4%)
Published	15 (50.0%)	6 (46.2%)	5 (50.0%)	4 (57.1%)

References

Bader MDM, Mooney SJ, Bennett B, & Rundle AG (2017). The promise, practicalities, and perils of virtually auditing neighborhoods using Google street view. The ANNALS of the American Academy of Political and Social Science, 669(1), 18–40. doi:10.1177/0002716216681488

Bader MDM, Mooney SJ, Lee YJ, Sheehan D, Neckerman KM,Rundle AG, & Teitler JO (2015). Development and deployment of the Computer Assisted Neighborhood Visual Assessment System (CANVAS) to measure health-related neighborhood conditions. Health & Place, 31, 163–172. doi: 10.1016/j.healthplace.2014.10.012 [PubMed: 25545769]

Barrett DP, Xu R, Yu H, & Siskind JM (2016). Collecting and annotating the large continuous action dataset. Machine Vision and Applications, 27(7), 983–995. doi:10.1007/s00138-016-0768-4

Bossard L, Guillaumin M, & Van Gool L (2014). Food-101 – Mining discriminative components with random forests In Fleet D, Pajdla T, Schiele B, & Tuytelaars T (Eds.), Computer vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, proceedings, part VI (pp. 446–461). Cham, Switzerland: Springer International Publishing.

Boushey CJ, Spoden M, Delp EJ, Zhu F, Bosch M, Ahmad Z, … Kerr D (2017). Reported energy intake accuracy compared to doubly labeled water and usability of the mobile food record among community dwelling adults. Nutrients, 9(3), 312. doi:10.3390/nu9030312

Boushey CJ, Spoden M, Zhu FM, Delp EJ, & Kerr DA (2016).New mobile methods for dietary assessment: Review of image- assisted and image-based dietary assessment methods. Proceedings of the Nutrition Society, 76(3), 283–294. doi:10.1017/S0029665116002913 [PubMed: 27938425]

Carlson JA, Liu B, Sallis JF, Kerr J, Hipp JA, Staggs VS, … Vasconcelos NM (2017). Automated ecological assessment of physical activity: Advancing direct observation. International Journal of Environmental Research and Public Health, 14(12), E1487. doi:10.3390/ijerph14121487 [PubMed: 29194358]

Colabianchi N (2015). Improving environmental measures in obesity research using innovative technology (1R21CA188481–01-A1). Washington, DC: NIH, Research Portfolio Online Reporting Tools (RePORT); Retrieved from https://projectreporter.nih.gov/.

Cowburn G, Matthews A, Doherty A, Hamilton A, Kelly P,Williams J, … Nelson M (2015). Exploring the opportunities for food and drink purchasing and consumption by teenagers during their journeys between home and school: A feasibility study using a novel method. Public Health Nutrition, 19(1), 93–103. doi:10.1017/S1368980015000889 [PubMed: 25874731]

Davis DS, Goldmon MV, & Coker-Appiah DS (2011). Using a community-based participatory research approach to develop a faith-based obesity intervention for african american children. Health Promotion Practice, 12(6), 811–822. doi: 10.1177/1524839910376162 [PubMed: 21540194]

Deng J, Dong W, Socher R, Li LJ, Kai L, & Li FF (2009,June 20–25). ImageNet: A large-scale hierarchical image database. Paper presented at the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL.

Evenson KR, Jones SA, Holliday KM, Cohen DA, & McKenzie TL (2016). Park characteristics, use, and physical activity: A review of studies using SOPARC (System for Observing Play and Recreation in Communities). Preventive Medicine, 86, 153–166. doi:10.1016/j.ypmed.2016.02.029 [PubMed: 26946365]

Eyler AA, Blanck HM, Gittelsohn J, Karpyn A, McKenzie TL,Partington S, … Winters M (2015). Physical activity and food environment assessments: Implications for practice. American Journal of Preventive Medicine, 48(5), 639–645. doi:10.1016/j.amepre.2014.10.008 [PubMed: 25891064]

Fulton JE, Carlson SA, Ainsworth BE, Berrigan D, Carlson C,Dorn JM, … Wendel A (2016). Strategic priorities for physical activity surveillance in the United States. Translational Journal of the American College of Sports Medicine, 1(13), 111–123. doi:10.1249/tjx.0000000000000020

Glanz K, Sallis JF, Saelens BE, & Frank LD (2007). Nutrition Environment Measures Survey in Stores (NEMS-S): Development and evaluation. American Journal of Preventive Medicine, 32(4), 282–289. doi:10.1016/j.amepre.2006.12.019 [PubMed: 17383559]

Graham DJ, & Hipp JA (2014). Emerging technologies to promote and evaluate physical activity: Cutting-edge research and future directions. Front Public Health, 2, 66. doi:10.3389/fpubh.2014.00066 [PubMed: 25019066]

Greene MR (2016). Estimations of object frequency are frequently overestimated. Cognition, 149, 6–10. doi:10.1016/j.cognition.2015.12.011 [PubMed: 26774103]

Hall KL, Vogel AL, Stipelman B, Stokols D, Morgan G, & Gehlert S(2012). A four-phase model of transdisciplinary team-based research: Goals, team processes, and strategies. Translational BehavioralMedicine, 2(4), 415–430. doi:10.1007/s13142-012-0167-y

Hipp JA, Adlakha D, Eyler AA, Chang B, & Pless R (2013). Emerging technologies: Webcams and crowd-sourcing to identify active transportation. American Journal of Preventive Medicine, 44(1), 96–97. doi:10.1016/j.amepre.2012.09.051 [PubMed: 23253658]

Hipp JA, Adlakha D, Eyler AA, Gernes R, Kargol A, Stylianou AH, & Pless R (2017). Learning from outdoor webcams: Surveillance of physical activity across environments In Thakuriah P,Tilahun N, & Zellner M (Eds.), Seeing cities through big data: Research, methods and applications in urban informatics (pp. 471–490). Cham, Switzerland: Springer International Publishing.

James P, Jankowska M, Marx C, Hart JE, Berrigan D, Kerr J, … Laden F (2016). "Spatial Energetics": Integrating data from GPS, accelerometry, and GIS to address obesity and inactivity. American Journal of Preventive Medicine, 51(5), 792–800. doi:10.1016/j.amepre.2016.06.006 [PubMed: 27528538]

Jenny Y, Russell B, Ce L, & Torralba A (2009, September 29–October 2). LabelMe video: Building a video database with human annotations. Paper presented at the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan.

Johansson G (1973). Visual perception of biological motion and a model for its analysis. Perception and Psychophysics, 14(2), 201–211. doi:10.3758/bf03212378

Joseph RP, & Maddock JE (2016). Observational park-based physical activity studies: A systematic review of the literature. Preventive Medicine, 89, 257–277. doi:10.1016/j.ypmed.2016.06.016 [PubMed: 27311337]

Kelly CM, Wilson JS, Baker EA, Miller DK, & Schootman M(2013). Using Google street view to audit the built environment: Inter-rater reliability results. Annals of Behavioral Medicine, 45(Suppl. 1), S108–S112. doi:10.1007/s12160-012-9419-9 [PubMed: 23054943]

Kelly P, Marshall SJ, Badland H, Kerr J, Oliver M, Doherty AR,& Foster C (2013). An ethical framework for automated, wearable cameras in health behavior research. American Journal of Preventive Medicine, 44(3), 314–319. doi:10.1016/j.amepre.2012.11.006 [PubMed: 23415131]

Kerr J, Duncan S, & Schipperjin J (2011). Using global positioning systems in health research: A practical approach to data collection and processing. American Journal of Preventive Medicine, 41(5), 532–540. doi:10.1016/j.amepre.2011.07.017 [PubMed: 22011426]

Kerr J, Marshall SJ, Godbole S, Chen J, Legge A, Doherty AR, … Foster C (2013). Using the SenseCam to improve classifications of sedentary behavior in free-living settings. American Journal of Preventive Medicine, 44(3), 290–296. doi: 10.1016/j.amepre.2012.11.004 [PubMed: 23415127]

King AC, Winter SJ, Sheats JL, Rosas LG, Buman MP, Salvo D, … Dommarco JR (2016). Leveraging citizen science and information technology for population physical activity promotion. Translational Journal of the American College of Sports Medicine, 1(4), 30–44. doi:10.1249/tjx.0000000000000003 [PubMed: 27525309]

Krizhevsky A, Sutskever I, & Hinton GE (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90. doi:10.1145/3065386

LeCun Y, Bengio Y, & Hinton G (2015). Deep learning. Nature, 521, 436. doi:10.1038/nature14539 [PubMed: 26017442]

Loveday A, Sherar BL, Sanders PJ, Sanderson WP, & Esliger WD (2015). Technologies that assess the location of physical activity and sedentary behavior: A systematic review. Journal of Medical Internet Research, 17(8), e192. doi:10.2196/jmir.4761 [PubMed: 26245157]

Macer PJ, Thomas PJ, Chalabi N, & Meech JF (1996). Finding the cut of the wrong trousers: Fast video search using automatic story- board generation Paper presented at the Conference Companion on Human Factors in Computing Systems, Vancouver, Canada.

Martin CK, Han H, Coulon SM, Allen HR, Champagne CM, & Anton SD (2009). A novel method to remotely measure food intake of free-living individuals in real time: The remote food photography method. British Journal of Nutrition, 101(3), 446–456. doi:10.1017/s0007114508027438 [PubMed: 18616837]

Martin D, Fowlkes C, Tal D, & Malik J (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics Paper presented at the Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, Vancouver, BC.

McKenzie TL, Cohen DA, Sehgal A, Williamson S, & Golinelli D(2006). System for Observing Play and Recreation in Communities (SOPARC): Reliability and feasibility measures. Journal of Physical Activity and Health, 3(s1), S208–S222. doi:10.1123/jpah.3.s1.s208

Medler DA (1998). A brief history of connectionism. Neural Computin gSurveys, 1, 61–101.

Moghimi M, Kerr J, Johnson E, Godbole S, & Belongie S (2015).Discriminative regions: A substrate for analyzing life-logging image sequences In He X, Luo S, Tao D, Xu C, Yang J, & Hasan MA (Eds.), MultiMedia Modeling: 21st International Conference, MMM 2015, Sydney, NSW, Australia, January 5–7, 2015, Proceedings, Part II (pp. 357–368). Cham, Switzerland: Springer International Publishing

Mooney SJ, Bader MDM, Lovasi GS, Teitler JO, Koenen KC,Aiello AE, … Rundle AG (2017). Street audits to measure neighborhood disorder: Virtual or in-person? American Journal of Epidemiology, 186(3), 265–273. doi:10.1093/aje/kwx004 [PubMed: 28899028]

Nahum-Shani I, Hekler EB, & Spruijt-Metz D (2015). Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association, 34(0), 1209–1219. doi:10.1037/hea0000306

National Institutes of Health. (2017). Certificates of Confidentiality (CoC).Retrieved from https://humansubjects.nih.gov/coc/index

Nebeker C, Lagare T, Takemoto M, Lewars B, Crist K, Bloss CS, & Kerr J (2016). Engaging research participants to inform the ethical conduct of mobile imaging, pervasive sensing, and location tracking research. Translational Behavioral Medicine, 6(4), 577–586. doi:10.1007/s13142-016-0426-4 [PubMed: 27688250]

Oliva A, & Torralba A (2006). Building the gist of a scene: The role of global image features in recognition. Progress in Brain Research, 155, 23–36. doi:10.1016/S0079-6123(06)55002-2 [PubMed: 17027377]

Park K, & Ewing R (2017). The usability of Unmanned Aerial Vehicles (UAVs) for measuring park-based physical activity. Landscape and Urban Planning, 167, 157–164. doi:10.1016/j.landurbplan.2017.06.010

Pless R, & Souvenir R (2009). A survey of manifold learning for images. IPSJ Transactions on Computer Vision and Applications, 1, 83–94. doi:10.2197/ipsjtcva.1.83

Ramanan D (2012). Detecting activities of daily living in first-person camera views Paper presented at the Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI.

Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, & Mermelstein R (2011). Health behavior models in the age of mobile interventions: are our theories up to the task? Translational Behavioral Medicine, 1(1), 53–71. doi:10.1007/s13142-011-0021-7 [PubMed: 21796270]

Rui Y, & Anandan P (2000). Segmenting visual actions based on spatiotemporal motion patterns Paper presented at the Proceedings IEEE Conference on Computer Vision and Pattern Recognition. (Cat. No. PR00662), Hilton Head Island, SC.

Spiegel PK (1995). The first clinical X-ray made in America–100 years.American Journal of Roentgenology, 164(1), 241–243. doi:10.2214/ajr.164.1.7998549 [PubMed: 7998549]

Streuber S, Quiros-Ramirez MA, Hill MQ, Hahn CA, Zuffi S,O'Toole A, & Black MJ (2016). Body talk: Crowdshaping realistic 3D avatars with words. ACM Translation Graph, 35(4), 1–14. doi:10.1145/2897824.2925981

Thornton IM, Rensink RA, & Shiffrar M (2002). Active versus passive processing of biological motion. Perception, 31(7), 837–853. doi:10.1068/p3072 [PubMed: 12206531]

Torralba A, Russell BC, & Yuen J (2010). LabelMe: Online image annotation and applications. Proceedings of the IEEE, 98(8), 1467–1484. doi:10.1109/JPROC.2010.2050290

U.S. Department of Health and Human Services. (2016). Child Welfare Information Gateway: Mandatory reporters of child abuse and neglect. Retrieved from https://www.childwelfare.gov/

Vaca-Castano G, Das S, Sousa JP, Lobo ND, & Shah M (2017).Improved scene identification and object detection on egocentric vision of daily activities. Computer Vision and Image Understanding, 156, 92–103. doi:10.1016/j.cviu.2016.10.016

Wang C, & Burris MA (1997). Photovoice: Concept, methodology, and use for participatory needs assessment. Health Education and Behavior, 24(3), 369–387. doi:10.1177/109019819702400309 [PubMed: 9158980]

Wang CC, & Pies CA (2004). Family, maternal, and child health through photovoice. Maternal and Child Health Journal, 8(2), 95–102. doi:10.1023/B:MACI.0000025732.32293.4f [PubMed: 15198177]

Whyte WH (1980). The social life of small urban spaces (2nd ed.).New York, NY: Project for Public Spaces.

Wood G, Lynch TP, Devine C, Keller K, & Figueira W (2016).High-resolution photo-mosaic time-series imagery for monitoring human use of an artificial reef. Ecology and Evolution, 6(19), 6963–6968. doi:10.1002/ece3.2342 [PubMed: 28725373]

Wood SA, Guerry AD, Silver JM, & Lacayo M (2013). Using social media to quantify nature-based tourism and recreation. Scientific Reports, 3, 2976. doi:10.1038/srep02976 [PubMed: 24131963]

Zacks JM, & Swallow KM (2007). Event segmentation. CurrentDirections in Psychological Science, 16(2), 80–84. doi:10.1111/j.1467-8721.2007.00480.x

Zhang Z, Hu Y, Chan S, & Chia LT (2008). Motion context: A new representation for human action recognition In Forsyth D, Torr P, & Zisserman A (Eds.), Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12–18, 2008, Proceedings, Part IV (pp. 817–829). Berlin, Heidelberg: Springer.

Zhuang Y, Belkin M, & Dennis S (2013). Metric based automatic event segmentation In Uhler D, Mehta K, & Wong JL (Eds.), Mobile Computing, Applications, and Services: 4th International Conference, MobiCASE 2012, Seattle, WA, USA, October 11–12, 2012. Revised Selected Papers (pp. 129–148). Berlin, Heidelberg: Springer.

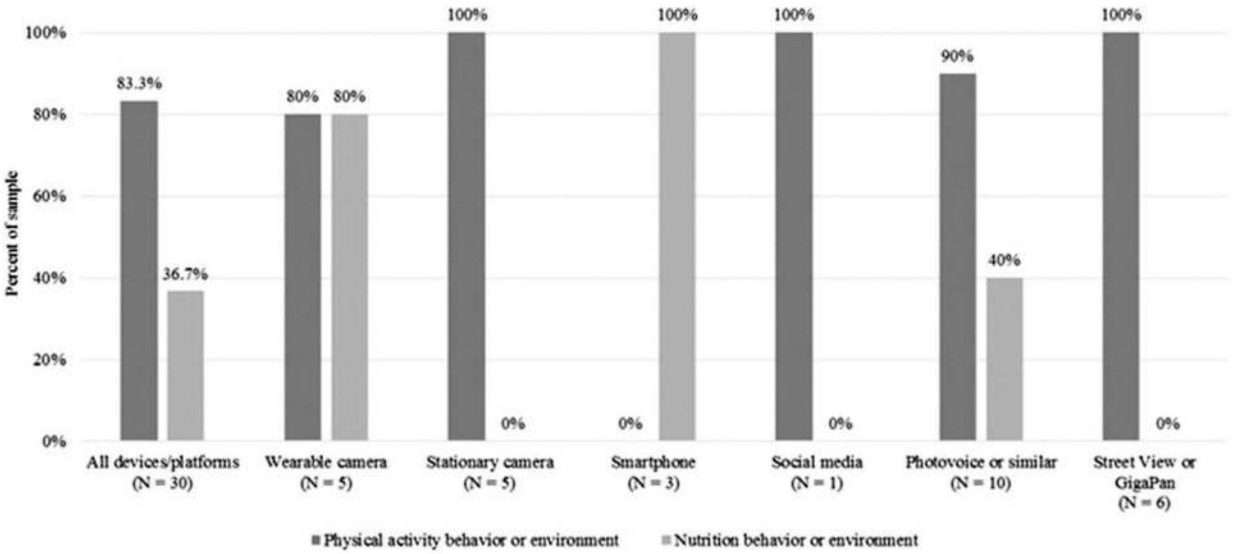
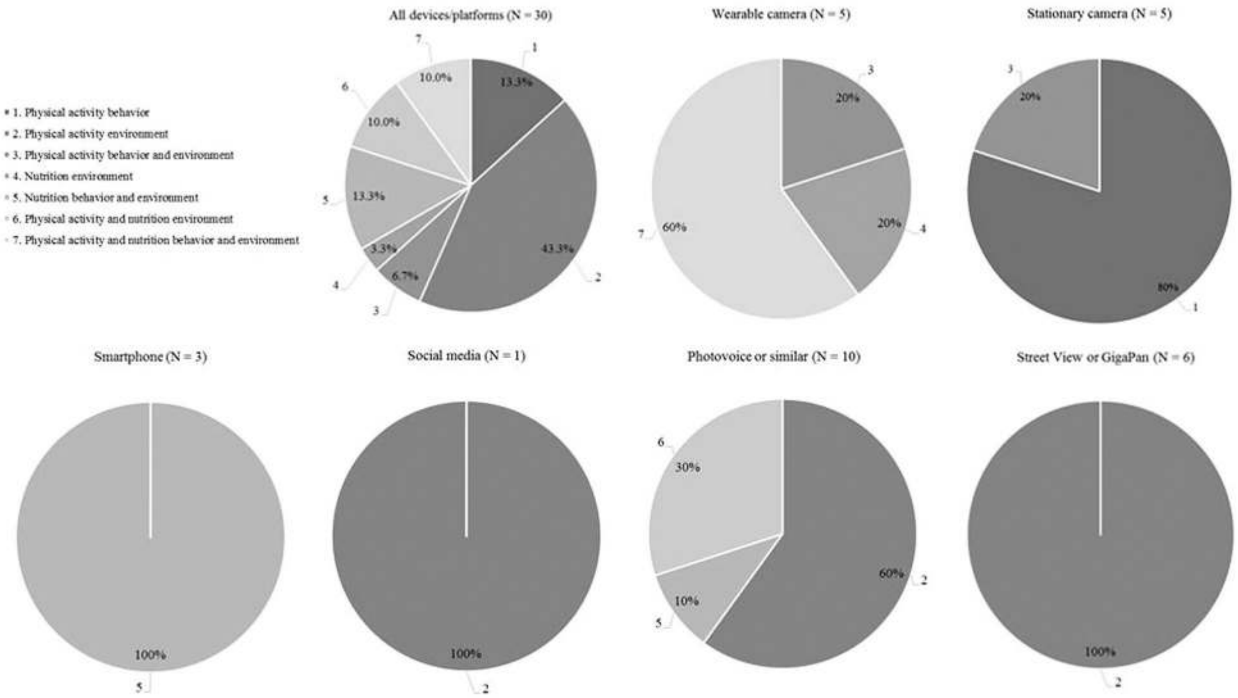


Figure 1—.

Coverage of physical activity and nutrition behaviors and environments by image capture device/platform used for items that were not mutually exclusive.



Percentages represent the percent of the sample endorsing the item
Items not shown were endorsed by 0% of the sample

Figure 2—.

Coverage of physical activity and nutrition behaviors and environments by image capture device/platform used for items that were mutually exclusive.

Table 1

Examples of Using Images in Research Related to Determinants of Obesity

Image Source	Behavior(s)	Environments	Source	Scoring	Citation
Gigapan	None	Street Segments and Parks	Researcher	Gigapan/Computer Automated	(Colabianchi, 2015)
Gigapan	Reef use	Coral Reefs	Researcher	CSIRO Ruggedized Autonomous Gigapixel System (CRAGS), for time series of high-resolution photo-mosaic (HRPM) imagery - Handstiched Panoramas and Manual coding	(Wood, Lynch, Devine, Keller, & Figueira, 2016)
Google Street View	None	Street Segments	Curated	Manual/89 Item Checklist	(Kelly, Wilson, Baker, Miller, & Schootman, 2013)
Google Street View	None	Street Segments/Neighborhood Disorder	Curated	Manual/Computer Assisted Neighborhood Visual Assessment System (CANVAS)	(Mooney et al.2017)
Photovoice	Safety, Engagement, Perception	Community Environment Assets	Participant	Group Discussion of Photos/Researcher Summary/Mapping of Sources	(Wang & Pies, 2004)
Photovoice (“Our Voice”)	Physical Activity	Neighborhood Social and Built Environment	Participant	Collective Review/Discussion, Healthy Neighborhood Discovery Tool	(King et al., 2016)
SenseCam	Sedentary and Transportation	Free-Living	Researcher	Semi-automated/Clarity Sensecam Browser	(Kerr et al., 2013)
SmartPhone	Dietary intake	Free-Living	Participant	Automated Food ID with user Confirmation	(Boushey et al., 2017)
SmartPhone	Food Selection/Plate Waste/ Energy intake	Free-Living	Participant	Manual/Comparison with Standard Photos	(Martin et al., 2009)
Social Media	Recreation Area Photo Posting	Parks/Recreation	Curated	Catalog of geotagged flickr images/Participant coding	(Wood, Guerry, Silver, & Lacayo, 2013)
Unmanned Aerial Vehicle (UAV) Mounted Video Camera	Activity	Parks	Researcher	Manual/Direct Observation of video	(Park & Ewing, 2017)
Video Camera	Use and Physical Activity	Parks and Schoolyards	Researcher	Computer Vision Algorithms: Features were extracted for each 1-s of video using deep convolutional neural networks for action recognition	(Carlson et al., 2017)
Webcam	Walking and Cycling	Street	Curated	Manual/Crowd Sourcing	(Hipp et al., 2013)

Note. Image source refers either to the technology that captured the image or the platform from which the image was obtained. Behavior(s) refers to the behavioral topic investigated in the research study. Environment(s) refers to the environment(s) investigated in the research study, with “free-living” indicating multiple environments. Source indicates who captured the images, with “curated” referring to researchers collecting existing images from third party sources. Scoring refers to how images were processed, with “Manual Scoring” referring to a human viewing each image and recording its features

Table 2

Challenges Using Image Capture in Research: Endorsement of Challenges Provided in Survey

		N (%) by Image Collection Category		
Research challenges (not mutually exclusive)	**Total *N* (%) (*N* = 30)**	**Participant Generated (*n* =13)**	**Researcher Generated (*n* =10)**	**Curated From Third Party (*n* = 7)**
Receptivity/recruitment	6 (20.0%)	3 (23.1%)	3 (30.0%)	0
Data collection	16 (53.3%)	7 (53.8%)	3 (30.0%)	6 (85.7%)
Data processing	19 (63.3%)	5 (38.5%)	10 (100%)	4 (57.1%)
Data quality	14 (46.7%)	6 (46.2%)	3 (30.0%)	5 (71.4%)
Ethical issues	10 (33.3%)	5 (38.5%)	4 (40.0%)	1 (14.3%)
Participant burden	6 (20.0%)	3 (23.1%)	3 (30.0%)	0
Scalability	12 (40.0%)	5 (38.5%)	6 (60.0%)	1 (14.3%)

Table 3

Challenges Using Image Capture in Research: Thematic Analysis of Open-Ended Qualitative Responses

		N (%) by Image Collection Category		
	Total *N* (%) (*N* =17)	**Participant Generated (*n* = 6)**	**Researcher Generated (*n* = 7)**	**Curated from Third Party (*n* = 4)**
Theme 1: Need for automated processing	10 (58.8%)	1(16.7%)	6 (85.7%)	3 (75.0%)
Theme 2: Quality of video/images	6 (35.3%)	3 (50.0%)	1 (14.3%)	2 (50.0%)
Theme 3: Successful collaborations	1 (5.9%)	0 (0%)	0 (0%)	1 (25.0%)
Theme 4: Participant recruitment/compliance	7 (41.2%)	4 (66.7%)	3 (42.9%)	0 (0%)
Theme 5: Device/platform challenges	4 (23.5%)	1 (16.7%)	1 (14.3%)	2 (50.0%)
Theme 6: Ethics/IRB	2 (11.8%)	1 (16.7%)	1 (14.3%)	0 (0%)
Theme 7: Scalability	4 (23.5%)	1(16.7%)	3 (42.9%)	0 (0%)

Table 4

Benefits of Using Image Capture in Research: Thematic Analysis of Open-Ended Qualitative Responses

		N (%) by Image Collection Category		
Theme and Brief Description	**Total *N* (%) (*N* = 26)**	**Participant Generated (*n* = 11)**	**Researcher Generated (*n* = 8)**	**Curated from Third Party (*n* = 7)**
Theme 1: Automates: Automated image processing was of value.	2 (7.7%)	0 (0%)	2 (25.0%)	0 (0%)
Theme 2: Reduces researcher burden/makes feasible: Images reduced the need for in-person observations.	6 (23.1%)	1 (9.1%)	1 (12.5%)	4 (57.1%)
Theme 3: Improves accuracy of measurement; Supports more objective assessment, reduces bias.	9 (34.6%)	1 (9.1%)	7 (87.5%)	1 (14.3%)
Theme 4: Potential for scalability: Involves an existing data source with wide coverage.	1 (3.8%)	0 (0%)	0 (0%)	1 (14.3%)
Theme 5: Provides more data: More settings and time points can be covered feasibly.	9 (34.6%)	2 (18.2%)	2 (25.0%)	5 (71.4%)
Theme 6: Supports evaluation: Provides existing data for evaluating naturalistic interventions.	2 (7.7%)	0 (0%)	1 (12.5%)	1 (14.3%)
Theme 7: Supports participant engagement/intervention: Participants are engaged through image capture.	8 (30.8%)	8 (72.7%)	0 (0%)	0 (0%)
Theme 8: Supports validation/algorithm training: Provides objective ground truth data.	2 (7.7%)	0 (0%)	2 (25.0%)	0 (0%)
Theme 9: Creates real-time or rapid feedback: Is being used to support real-time analyses and actuations.	1 (3.8%)	0 (0%)	1 (12.5%)	0 (0%)
Theme 10: Generates qualitative data to complement quantitative: Supports improved contextual information through imagery and descriptive summaries.	6 (23.1%)	6 (54.5%)	0 (0%)	0 (0%)