The Smartphone As a Platform for Wearable Cameras in Health Research

Cathal Gurrin, PhD, Zhengwei Qiu, MSc, Mark Hughes, PhD, Niamh Caprani, MSc,
Aiden R. Doherty, PhD, Steve E. Hodges, PhD, Alan F. Smeaton, PhD

Background: The Microsoft SenseCam, a small camera that is worn on the chest via a lanyard, increasingly is being deployed in health research. However, the SenseCam and other wearable cameras are not yet in widespread use because of a variety of factors. It is proposed that the ubiquitous smartphones can provide a more accessible alternative to SenseCam and similar devices.

Purpose: To perform an initial evaluation of the potential of smartphones to become an alternative to a wearable camera such as the SenseCam.

Methods: In 2012, adults were supplied with a smartphone, which they wore on a lanyard, that ran life-logging software. Participants wore the smartphone for up to 1 day and the resulting life-log data were both manually annotated and automatically analyzed for the presence of visual concepts. The results were compared to prior work using the SenseCam.

Results: In total, 166,000 smartphone photos were gathered from 47 individuals, along with associated sensor readings. The average time spent wearing the device across all users was 5 hours 39 minutes (SD=4 hours 11 minutes). A subset of 36,698 photos was selected for manual annotation by five researchers. Software analysis of these photos supports the automatic identification of activities to a similar level of accuracy as for SenseCam images in a previous study.

Conclusions: Many aspects of the functionality of a SenseCam largely can be replicated, and in some cases enhanced, by the ubiquitous smartphone platform. This makes smartphones good candidates for a new generation of wearable sensing devices in health research, because of their widespread use across many populations. It is envisioned that smartphones will provide a compelling alternative to the dedicated SenseCam hardware for a number of users and application areas. This will be achieved by integrating new types of sensor data, leveraging the smartphone's real-time connectivity and rich user interface, and providing support for a range of relatively sophisticated applications.
(Am J Prev Med 2013;44(3):308–313) © 2013 American Journal of Preventive Medicine

Introduction

As highlighted elsewhere[1–5] in this special issue, the Microsoft SenseCam[6] increasingly is being deployed in a preventive medicine context and early results are promising. However, despite its potential, it is not yet in widespread use at the population level. One hurdle to adoption of the SenseCam may be the necessity for users to purchase, maintain, and operate a dedicated hardware device. Although the benefits for some users may make this effort worthwhile,[7] it is proposed that adoption at scale would be facilitated greatly if the SenseCam functionality largely could be integrated with a device that users already own and are accustomed to charging and maintaining. An obvious avenue is the increasingly ubiquitous smartphone.

Previous research has shown that smartphones can be used as a platform for lifestyle monitoring.[8,9] However, until recently, smartphones have not been able to passively sample a broad range of lifestyle health-related behaviors as effectively as SenseCam because of a number of issues. These include battery lifetime while capturing photos, limited built-in sensing, limited field-of-view, and form factor.

In particular, using a smartphone for sampling life activities naturally poses a challenge in terms of battery usage.[10] To be an effective life-sampling tool, a smartphone needs to operate all day without requiring recharg-

ing, in a manner similar to the SenseCam. Fortunately, a number of recent explorations of power-efficient sensing and communication with smartphones can be leveraged to provide all-day operation. Examples include the ACQUA (Acquisition Cost-Aware Query Adaption) framework,[11] power-optimized scheduling of data transfer,[12] and the life-sampling tool of Qiu et al.[13]

There potentially would be several benefits to using a smartphone as a SenseCam replacement. Automatic upload of sensed data to a cloud-based server for processing could open up a number of possibilities for real-time data analysis and intervention. The smartphone's networking hardware also could be used to record the user's proximity to WiFi networks and other users' Bluetooth devices throughout the day. Such data could provide valuable information on the whereabouts of the wearer and their interactions with other people.[14] GPS sensing, now commonplace in smartphones, provides additional location information that could be utilized.

In this work, a smartphone application was developed that operates in a manner similar to a SenseCam, and it is shown that the smartphone can capture accurate and meaningful life activities over a full day. Data were sampled constantly from the full gamut of built-in sensors: accelerometer, compass, camera, GPS, Bluetooth, microphone and WiFi, as well as from external sensors via Bluetooth.[13] Although some SenseCam sensing modalities such as temperature and light level are missing, for many applications the smartphone provides a substantially richer sensor set than the SenseCam. It also has the ability to support a larger suite of lifestyle and behavior analysis tools.

Although the SenseCam was designed especially as a wearable device, the smartphone is, of course, primarily a communications device optimized for handheld interaction. However, with the gradual size and weight reduction of smartphones, the authors believe that in many cases the smartphone can approach the "wearability" of SenseCam. Figure 1 shows a comparison between an individual wearing a SenseCam (left) and a smartphone (right).

A wearable camera/sensor platform, such as the smartphone implementation that is envisaged in this work, must be able to gather data from all on-board sensors continually during the day. A secondary aim is to automatically identify specified daily activities using the passively captured image and sensor data. To this end, this article addresses the following two research questions:

1. How much wear-time can be achieved using smartphones intended to mimic SenseCam functionality?

Figure 1. An individual wearing a Microsoft SenseCam (left) and a smartphone (right)

2. Can the automated detection of lifestyle-related activities from smartphones be as accurate as that from the SenseCam?

It is expected that the results of the present study will be of benefit to researchers who want to exploit the ubiquity of smartphones to investigate human behavior. In particular, the aim here is to enable researchers to leverage smartphones to automatically capture lifestyle behaviors.

Methods

Recruitment of Participants

To evaluate the potential of smartphones as wearable cameras, 47 participants were recruited in 2012. These users were chosen from a broad spectrum of user groups, from students to retired individuals, with varying incomes, age ranges, and education levels. Each participant gave consent for this study, and all study procedures were approved by the research ethics board of Dublin City University.

Each participant was asked to wear the smartphone for up to 8 hours in a given day, so as to capture an entire range of everyday activities. The smartphone was worn on a lanyard around the neck, as is the case with a SenseCam (Figure 1). Participants were given the phone with the software already installed and ready to run; the phone was returned after the data-gathering session was complete. The smartphone was provided in addition to the participant's normal phone; thus, it was not used as a phone during the experiment.

Data Collection

The smartphone includes sensors that can capture a rich life-experience archive. Using an Android (OS version 2.1) mobile phone, life-logging software was developed that captured photos at a configurable rate along with data gathering from all the available on-board sensors. Unlike with the SenseCam, automatic analysis of the user context was carried out in real time directly on the smartphone. This allows a power-efficient sampling algorithm to be used, frequently resulting in a full day's operation.[11]

Smart upload of content means that a minimal representation of a new activity—important image and sensor data—can be streamed to a server. The remaining data are uploaded via WiFi to a remote server when charging the device. The server contains a database to store references to the image and sensor data. In addition, the server stores the images, which are then available for viewing from any PC/TV/mobile device. For more details of implementation, please refer to Qiu et al.[13]

Table 1. Distribution and density of daily activities with the accuracy of performance of concept detectors

Daily activities	User distribution (%)	Density (%)	Precision	Recall	Daily activities	User distribution (%)	Density (%)	Precision	Recall
Bar (interior)	26	2.5	0.90	0.70	Outdoor	83	19.3	0.83	0.86
Bottle	64	1.2	0.74	0.59	Person	85	6.6	0.66	0.65
Bus (exterior)	57	0.4	0.64	0.61	Road/path	81	10.7	0.82	0.86
Bus (interior)	19	1.2	0.69	0.55	Shop (interior)	62	4.2	0.83	0.86
Cafe/restaurant (interior)	17	1.7	0.60	0.33	Shopping center (interior)	36	0.7	0.72	0.73
Car (interior)	55	16.9	0.88	0.83	Small/mobile digital screen (handheld)	38	0.5	0.66	0.53
Crowd	11	0.3	0.68	0.49	Stage	9	0.4	0.76	0.74
Cup	30	0.4	0.77	0.42	Steering wheel	40	8.0	0.86	0.79
Food	32	0.4	0.78	0.50	Train (interior)	9	0.5	0.92	0.52
Glass	43	1.7	0.70	0.43	TV/computer screen	47	5.4	0.81	0.41
Hand	81	7.7	0.67	0.38	Written media (book/magazine/ newspaper)	47	0.8	0.81	0.68
Indoor	81	37.8	0.87	0.82	Eating/snacking	47	1.8	0.68	0.61

Note: For each visually identifiable concept associated with an activity, the User Distribution refers to the percentage of users that encountered that activity and Density refers to the percentage of the image collection that contained that concept. Precision and Recall are generally accepted measures of accuracy.

To evaluate the effectiveness of the smartphone as a wearable sensing tool for lifestyle behaviors, participants were supplied with the smartphone configured to automatically capture location-stamped photos every 6 seconds (around three to four times the SenseCam rate, but this is configurable). Other sensors were sampled at the following rates: accelerometer 5 times per second, compass 10 times per minute, GPS every 20 meters, Bluetooth and WiFi every 2 minutes, microphone once per minute, and communication/media activities as they occurred. Participants were asked to wear the smartphone for up to 8 hours in the day, during 2012, and were asked to go about their everyday free-living activities. This period has been used in prior accelerometer-based studies[15] and is comparable to the usage of SenseCam in many cases.[16]

Automated Annotation of Daily Activities

Certain objects or concepts inherent in smartphone images can be classified automatically by visually analyzing the photo streams, with or without the mining of additional sensors. In this way, it is possible to identify activities that the user is involved in. For example, the presence of cups and plates in a photo stream can indicate that the wearer is eating.

For this experiment, all photos were processed automatically using well-developed computer vision techniques based on speeded-up robust features (SURF)[17] combined with the raw sensor streams (e.g., GPS). The colors of each image (ColorStructure and ScalableColor from the MPEG-7 computer vision libraries[18]) also were extracted and used as input for automatic identification of visual objects using an optimized SVMlight implementation of the Support Vector Machine[19] with crossfold validation. This image analysis technique commonly is used in the image and video search domains.[20]

Selection and Coding of Visual Concepts

An exemplar set of 24 visual concepts was selected for initial exploration. These concepts, and their frequency of identification by users and in aggregate, are shown in Table 1. For each feature, the User Distribution refers to the percentage of users that encountered that visual concept, and density refers to the percentage of the image collection that contained that concept. Using a subsample of temporally selected photos, each image was annotated for the presence of each of the concepts listed in Table 1. To ensure accuracy, each photo was annotated by multiple annotators.

Data Analysis

Summary statistics on the number of images and total wear time across all users were gathered. The software parameter tuning for each lifestyle concept was trained using the annotated training images. The performance of concept identification was then evaluated on the annotated test collection. For this evaluation, k-fold cross-validation was used. This is a technique in which the training set is split into k disjoint subsets of equal sizes and a model is trained for each subset, with the overall performance of the classifier being calculated as the mean accuracy of each subset. In this way, the optimization of the software for a single set of training and test data can be avoided.

In this work, a value of 5 is chosen for k, which has been determined empirically to provide a reasonable balance between processing time required to train each model and the accuracy of the validation. In addition, the data were subsampled to ensure that the same number of positive and negative examples were used when training and evaluating the models, so as to eliminate bias in the

training. As this is a machine-vision classification challenge, standard information retrieval metrics were used: precision (a positive predictive value); recall (same as sensitivity); and F1 score, a single value that reflects both precision and recall.

Results

1. How much wear-time can be achieved using smartphones that mimic SenseCam functionality?

In total, 166,000 photos were gathered from the 47 individuals, along with the associated GPS readings during the experiment in 2012. The average time spent wearing the device across all users was 5 hours 39 minutes (SD=4 hours 11 minutes), ranging from 11 minutes to more than 18.5 hours. The higher end of the scale of time spent wearing the device clearly shows that under the right conditions the smartphone can last for a whole day on a single charge. This is discussed further in the Discussion section.

2. Can the automated detection of lifestyle-related activities from smartphones be as accurate as that from the SenseCam?

A subset of 36,698 photos were selected for manual annotation by five researchers in 2012. A total of 873,552 annotation judgments were carried out on this subset. The results from the annotation process are shown in Table 1. Thereafter, the subset of photos was processed automatically using state-of-the-art computer-vision techniques, producing the set of life activity–detection accuracy figures.

Discussion

This experiment has shown that it is possible to replicate the functionality of a passive image-capture device (such as the SenseCam) on a smartphone and that smartphones provide an effective platform for implementing wearable-camera technology. The current study has found that across 47 participants, the average time to wear the device was 5 hours 39 minutes, which is less than that observed in other SenseCam-related studies (e.g., 8 hours 45 minutes in Doherty et al.[16]). The large SD of 4 hours 11 minutes appears to be due to both user and technical issues. Some participants may have recharged their device in the middle of the day; thus, a stricter protocol should be developed for future studies.

In addition, there are variations in power consumption that occur as a result of the real-time sensor data processing. For example, the power-hungry GPS sensor is only used if other sensors such as WiFi and Bluetooth are unable to detect location. It is expected that disabling location sensing would reduce this variation substantially and result in consistently longer operating times. In this study, the smartphones were configured to capture an image every 5 seconds, in contrast to the rate of once every 15–20 seconds for the SenseCam.[6] Setting a lower capture rate would further increase battery lifetime.

The automated machine-vision technique used to analyze the images was able to identify a range of lifestyle concepts with an average F1 score of 0.67. Although they are on a slightly different set of visual concepts, the results of this Android-based smartphone study are very similar to previously reported identification of SenseCam images. Using SenseCam images, a mean F1 score of 0.65 was calculated on the basis of a set of 95,000 manually annotated images.[16] This shows comparable performance for the two devices as wearable-camera platforms.

Smartphones are a suitable lifestyle and behavioral concept monitoring device for the following reasons:

1. All-day operation is possible if images and other sensor-logging is not done too aggressively.
2. The presence of a full suite of on-board sensors such as the accelerometer, compass, camera, GPS, Bluetooth, microphone, WiFi and communication/media activity sensors, as well as the ability to sense from external sensors via Bluetooth can be useful for many applications.
3. The ubiquity, cost, and familiarity of the smartphone are better compared to those for a SenseCam.
4. The smartphone can support real-time analysis of sampled life experiences, enabling applications requiring immediate intervention or prospective memory cues.
5. The ability to act as both the data-gathering device and the data-display and feedback device, allowing for ubiquitous access to the sampled life activities, are an essential aspect of supporting real-time health behavior interventions.
6. The ability to update the software to support new methods of sampling or interaction in a proven and easy-to-achieve manner is valuable.

Limitations

This work has highlighted some limitations associated with using a smartphone for visual life-logging. One slight drawback is that some current smartphones are slightly bulkier than the SenseCam, which may have ergonomic implications. In addition, typical smartphones are not equipped with a fish-eye (wide-angle) lens to capture a wide field of view, whereas a SenseCam does have such a lens.

For activities that typically happen in front of the individual, this should not pose a problem, although for more general life-logging applications, a narrow (conventional) field of view is not ideal. However, it is possible to buy and fit a fish-eye lens onto a smartphone to offset this limitation. The orientation of the smartphone camera typically results in a portrait view, although the addition of a fish-eye lens will help to offset any negative impact of this narrower view.

Figure 2 provides a comparison of the fish-eye lens on a SenseCam (left) with the normal field-of-view lens of a smartphone (right).

The choice was made to limit the study to a single day for each participant because it was felt that this would generate suitably representative data for an initial analysis. Having validated the basic principle of a worn smartphone as a SenseCam replacement, the authors would like to refine understanding of the relative merits of the two technologies through further studies, gathering evidence across a wider range of activities by way of more users and longer periods of use. In addition, the sample of volunteer users had to carry an additional smartphone at all times and in future studies, the authors would like to move to a single-smartphone solution in order to evaluate any issues this introduces; it is felt that this is the most realistic scenario for a long-term, real-world deployment.

The 24-exemplar visual concepts selected for analysis in this article do not necessarily reflect an optimum selection of health-related behaviors. They were selected in part on the basis of saliency for robust analysis. Although they were appropriate for the purposes of an exploratory study to investigate whether the current method can be a tool for behavioral scientists, the authors plan to develop an improved set in future work. The learning process for any newly selected concepts is the same as that for the 24 used in this exploration.

Conclusion

Wearable cameras can provide invaluable information relating to the behavior of an individual throughout the course of the day. This work demonstrates that smartphones can now be considered an alternative platform to dedicated devices such as the SenseCam for future studies. Although smartphones have some drawbacks, they also exhibit a number of advantages when compared to SenseCam. This will be beneficial for a number of researchers working in this domain. In particular, future preventive medicine research, both observation and intervention, may leverage wearable cameras using the smart-

Figure 2. Comparing a Microsoft SenseCam with fish-eye lens (left) and a smartphone with a conventional lens (right)
Note: Take note of the absence of the individual on the right from the smartphone photo.

phone. In the longer term, platforms such as smartphones will make it more feasible to use wearable camera technologies in a wider range of populations.

References

1. Kerr J, Marshall SJ, Godbole S, et al. Using the SenseCam to improve classifications of sedentary behavior in free-living settings. Am J Prev Med 2013;44(3):290–296.
2. O'Laughlin G, Cullen SJ, McGoldrick A, et al. Using a wearable camera to increase the accuracy of dietary analysis. Am J Prev Med 2013;44(3):297–301.
3. Silva AR, Pinho S, Macedo LM, Moulin CJ. Benefits of SenseCam review on neuropsychological test performance. Am J Prev Med 2013;44(3):302–307.
4. Kelly P, Badland H, Kerr J, et al. An ethical framework for automated, wearable cameras in health behavior research. Am J Prev Med 2013;44(3):314–319.
5. Doherty AR, Hodges SE, King A, et al. Wearable cameras in health: the state of the art and future possibilities. Am J Prev Med 2013;44(3):320–323.
6. Hodges S, Williams L, Berry E, et al. SenseCam: a retrospective memory aid. Proceedings of Ubicomp 2006, LNCS 4206:177–93.
7. Berry E, Hampshire A, Rowe J, et al. The neural basis of effective memory therapy in a patient with limbic encephalitis. J Neurol Neurosurg Psychiatry 2009;80(3):582–601.

8. Doherty AR, Tolle KM, Smeaton AF. Utilising contextual memory retrieval cues and the ubiquity of the cell phone to review lifelogged physiological activities. Proceedings of the IMCE 2009—International Workshop on Interactive Multimedia for Consumer Electronics; 2009 Oct 23; Beijing, China.

9. Dunton GF, Liao Y, Intille S, Spruijt-Metz D, Pentz M. Investigating children's physical activity and sedentary behavior using ecological momentary assessment with mobile phones. Obesity (Silver Spring) 2011;19(6):1205–12.

10. Arab L, Estrin D, Kim DH, Burke J, Goldman J. Feasibility testing of an automated image-capture method to aid dietary recall. Eur J Clin Nutr 2011;65(10):1156–62.

11. Misra A, Lim L. Optimizing sensor data acquisition for energy-efficient smartphone-based continuous event processing. Proceedings: Mobile Data Management (1) 2011:88–97.

12. Dogar FR, Steenkiste P, Papagiannaki K. Catnap: exploiting high bandwidth wireless interfaces to save energy for mobile devices. Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services (MobiSys '10). ACM; 2010; New York NY, 107–22.

13. Qiu Z, Gurrin C, Doherty A, Smeaton AF. A real-time life experience logging tool. Proceedings of the 18th Multimedia Modelling Conference (MMM 2012): Advances in Multimedia Modeling, 636–638, 2012.

14. Eagle N, Pentland AS, Lazer D. Inferring friendship network structure by using mobile phone data. Proc Natl Acad Sci U S A 2009; 106(36):15274–8.

15. Ruiz JR, Ortega FB, Martínez-Gómez D, et al. Objectively measured physical activity and sedentary time in European adolescents. Am J Epidemiol 2011;174(2):173–84.

16. Doherty AR, Caprani N, Conaire CO, et al. Passively recognising human activities through lifelogging. Comput Human Behav 2011; 27(5):1948–58.

17. Bay H, Tuytelaars T, Van Gool L. Surf: speeded up robust features. Proceedings of the 9th European Conference on Computer Vision, 2006.

18. Sikora T. The MPEG-7 visual standard for content description—An overview. IEEE Trans Circuits Syst Video Technol 2001;11(6):696–702.

19. Joachims T. Training linear SVMs in linear time. Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06), ACM; 2006; New York, NY.

20. Snoek CGM, Worring M. Concept-based video retrieval. Found Trends Inform Retriev 2009;2(4):215–322.