

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

From: CLARITY Centre for Sensor Web Technologies (Gurrin, Qiu, Hughes, Caprani, Doherty, Smeaton), Dublin City University, Ireland,

Microsoft Research Cambridge (Hodges), United Kingdom

Address correspondence to: Cathal Gurrin, PhD, School of Computing, Dublin City University, Glasnevin, Dublin 9, Ireland. E-mail: cgurrin@computing.dcu.ie.

Background: The SenseCam, a small camera that is worn on the chest via a lanyard, is increasingly being deployed in a health research. However, the SenseCam is not yet in widespread use due to a variety of factors. It is proposed that the ubiquitous smartphone can provide a more accessible alternative to the SenseCam.

Purpose: To perform an initial evaluation of the potential of the smartphone to become an alternative to the SenseCam.

Methods: Adults were supplied with a smartphone running lifelogging software which they wore on a lanyard. Participants wore the smartphone for up to a day and the resulting lifelog data was 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 and 39 minutes (\pm 4 hours 11 minutes SD). A subset of 36,698 photos was selected for manual annotation by 5 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 can largely be replicated, and in some cases enhanced, by the ubiquitous smartphone platform. This makes the smartphone a good candidate for a new generation of wearable sensing devices in health research, due to its widespread use across many populations. It is envisioned that the smartphone 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.

Introduction

As highlighted elsewhere in this special issue, the SenseCam¹ is being increasingly deployed in a preventive medicine context and early results are promising. However, despite its potential it is not yet in significant use at the population level. One hurdle to the widespread adoption of the SenseCam may be the necessity for users to purchase, maintain and operate a dedicated hardware device. While the benefits for some users may make this effort worthwhile,² it is proposed that adoption at scale would be greatly facilitated if the SenseCam functionality could largely 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.^{3,4} 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 such as: 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 for battery usage.⁵ To be an effective life sampling tool, a smartphone needs to operate all day without requiring recharging, 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. For example, the ACQUA (Acquisition Cost-Aware Query Adaption) framework,⁶ power optimized scheduling of data transfer⁷ and the life sampling tool of Qiu et al.⁸

There would potentially 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 could also 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.⁹ 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 was 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.⁸ Although some SenseCam sensing modalities such as temperature and light level are missing, for many applications the smartphone provides a significantly richer sensor set than the SenseCam. It also has the ability to support a larger suite of lifestyle and behavior analysis tools. While the SenseCam was especially designed as a wearable device, the smartphone is of course primarily a communications device optimised for handheld interaction. However, with the gradual size and weight reduction of smartphones we 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 the 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?
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 this study will be of benefit to researchers who want to exploit the ubiquity of smartphones to investigate human behaviour. In particular, we aim to enable researchers to leverage smartphones to automatically capture lifestyle behaviours.

Methods

Recruitment of Participants

To evaluate the potential of the smartphone as a wearable visual lifelogging tool, 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, which was approved by the DCU ethics committee. Each participant was asked to wear the smartphone for up to eight 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 (see 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 and as such 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, lifelogging software was developed that captured photos at a configurable rate along with data gathering from all the available on-board sensors. Unlike 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 operation⁶. 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 is 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 view from any PC / TV / mobile device. For more details of implementation, please refer to Qiu et al.⁸

In order to evaluate the effectiveness of the smartphone as a wearable sensing tool for lifestyle behaviours, participants were supplied with the smartphone configured to automatically capture location-stamped photos every six seconds (around 3-4 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 metres, 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 eight 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 studies¹⁰ and is comparable to the usage of SenseCam in many cases¹¹.

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 automatically processed using well-developed computer vision techniques based on speeded up robust features (SURF) features¹² combined with the raw sensor streams (e.g. GPS). The colors of each image (ColorStructure and ScalableColor from the MPEG-7 computer vision libraries¹³) were also extracted and used as input for automatic identification of visual objects using an optimised SVMlight implementation of the Support Vector Machine¹⁴ with cross-fold validation. This image analysis technique is commonly used in the image and video search domains.¹⁵

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, is 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.

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 where 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 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 empirically determined to provide a reasonable balance between processing time required to train each model and the accuracy of the validation. In addition, the data was 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 and 39 minutes (± 4 hours 11 minutes SD), ranging from 11 minutes to over 18½ 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 next 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 5 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 were automatically processed 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 the smartphone provides an effective platform for implementing wearable camera technology. This 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.¹¹). The large standard deviation 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, therefore in future studies a more strict protocol would need to be developed. In addition there are variations in power consumption which 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 SenseCam¹. Setting a lower capture rate would further increase battery lifetime.

The automated machine vision technique used to analyse the images were able to identify a range of lifestyle concepts with an average F1 score of 0.67. While 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 images¹¹. This shows a comparable performance between both devices as wearable camera platforms.

It is concluded that the Smartphone is a suitable lifestyle and behavioral concept monitoring device for the following reasons:

- All-day operation is possible if images and other sensor logging is not done too aggressively.
- 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.
- The ubiquity, cost and familiarity of the smartphone are better when compared with a SenseCam.
- The smartphone can support real-time analysis of sampled life experiences, enabling applications requiring immediate intervention or prospective memory cues.
- The ability to act as both the data gathering device, as well as 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.

- 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 lifelogging. One slight drawback is that some current smartphones are slightly bulkier than the SenseCam, which may have ergonomic implications. Also typical smartphones are not equipped with a fish-eye (wide-angle) lens to capture a wide field of view, while a SenseCam does have such a lens. For activities which typically happen in front of the individual, this should not pose a problem, though for more general lifelogging applications, a narrower (conventional) field of view is not ideal. However it is possible to buy and fit a fisheye lens onto a smartphone to offset this limitation. The orientation of the smartphone camera typically results in a portrait view, though 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 fisheye lens on a SenseCam (left) with the normal field-of-view lens of a smartphone (right).

It was chosen 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, we would like to refine our 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 we would like to move to a single-smartphone solution in order to evaluate any issues which 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 behaviours. They were selected in part on the basis of saliency for robust analysis. While they were appropriate for the purposes of an exploratory study to investigate if our method can be a tool for behavioural scientists, we plan to develop an improved set in future work. The learning process for any newly selected concepts is the same as for the 24 we use in this exploration.

Conclusion

We believe that wearable cameras can provide invaluable information relating to the behavior of an individual throughout the course of the day. We hope that this work demonstrates to others working in the field that smartphones can now be considered as 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. We think 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 visual lifelogging using the smartphone. In the longer term, platforms like the smartphone will make it more feasible to use lifelogging technologies in a wider range of populations.

Acknowledgements

Author AD is a Marie Curie postdoctoral research fellow supported by the Irish Health Research Board under grant number MCPD/2010/12. Authors AD and AS are supported by the Science Foundation Ireland under Grant No. 07/CE/I1147.

No financial disclosures were reported by the authors of this paper.

References

1. Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., and Wood, K. (2006). SenseCam: a Retrospective Memory Aid. In Proceedings of Ubicomp 2006, LNCS 4206, pp. 177 – 193.
2. Berry, E., Hampshire, A., Rowe, J., Hodges, S., Kapur, N., Watson, P., Smyth, G. B. G., et al. (2009). The neural basis of effective memory therapy in a patient with limbic encephalitis. *Neurology, Neurosurgery, and Psychiatry with Practical Neurology*, 80(3), 582–601.
3. Doherty A R, Tolle K M, Smeaton A F. Utilising contextual memory retrieval cues and the ubiquity of the cell phone to review lifelogged physiological activities. In: IMCE 2009 - International Workshop on Interactive Multimedia for Consumer Electronics, 23 October 2009, Beijing, China.
4. Dunton GF, Liao Y, Intille S 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):1205-1212.
5. Arab L, Estrin D, Kim D H, Burke J, Goldman J. Feasibility testing of an automated image-capture method to aid dietary recall. *Eur J Clin Nutr.* 2011 Oct;65(10):1156-62. doi: 10.1038/ejcn.2011.75. Epub 2011 May 18.
6. Misra A, Lim L. Optimizing Sensor Data Acquisition for Energy-Efficient Smartphone-Based Continuous Event Processing. *Mobile Data Management (1)* 2011: 88-97
7. Dogar F R, Steenkiste P, Papagiannaki K. Catnap: exploiting high bandwidth wireless interfaces to save energy for mobile devices. In *Proceedings of the 8th international conference on Mobile systems, applications, and services (MobiSys '10)*. ACM, New York, NY, USA, 107-122.

8. Qiu Z, Gurrin C, Doherty A, Smeaton AF. A Real-Time life experience logging tool. 18th Multimedia Modelling Conference (MMM 2012). *Advances in Multimedia Modeling*, pages 636–638, 2012.
9. Eagle, N., Pentland, A. S., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America*, 106(36), 15274-15278.
10. Ruiz J R, Ortega F B, Martínez-Gómez D, Labayen I, Moreno L A, De Bourdeaudhuij I, Manios Y, Gonzalez-Gross M, Mauro D, Molnar D, Widhalm K, Marcos A, Beghin L, Castillo M J, Sjöström M (on behalf of the HELENA Study Group). Objectively Measured Physical Activity and Sedentary Time in European Adolescents. 2011/04/05. *American Journal of Epidemiology*. 10.1093/aje/kwr068
11. Doherty A R, Caprani N, Conaire C O, Kalnikaite V, Gurrin C, Smeaton AF, O'Connor N E. Passively recognising human activities through lifelogging. *Computers in Human Behavior* 2011, 27(5):1948-1958.
12. Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features. *9th European Conference on Computer Vision, 2006*.
13. Sikora T, The MPEG-7 visual standard for content description-an overview, *Circuits and Systems for Video Technology, IEEE Transactions on* , vol.11, no.6, pp.696-702, Jun 2001
14. Joachims T. Training linear SVMs in linear time. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '06)*. ACM, New York, NY, USA,
15. Snoek C G M, Worring M. Concept-Based Video Retrieval. *Foundations and Trends® in Information Retrieval: Vol. 2: No 4*, pp 215-322.

Figure Legends

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

Figure 2. Comparing a SenseCam with fisheye lens (left) and a smartphone with a conventional lens (right). Note the absence of the individual on the right from the smartphone photo.

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 centre (Interior)	36%	0.7%	0.72	0.73
Car (interior)	55%	16.9%	0.88	0.83	Small/mobile digital screen (hand-held)	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/Mag/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

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.