Mining User Activity as a Context Source for Search and Retrieval

Zhengwei Qiu,Aiden R. Doherty, Cathal Gurrin, Alan F. Smeaton
CLARITY: Centre for Sensor Web Technologies, School of Computing, Dublin City University, Ireland
{zqui, adoherty,cgurrin, asmeaton}@computing.dcu.ie

***Abstract*—Nowadays in information retrieval it is generally accepted that if we can better understand the context of searchers then this could help the search process, either at indexing time by including more metadata or at retrieval time by better modelling the user needs. In this work we explore how activity recognition from tri-axial accelerometers can be employed to model a user's activity as a means of enabling context-aware information retrieval. In this paper we discuss how we can gather user activity automatically as a context source from a wearable mobile device and we evaluate the accuracy of our proposed user activity recognition algorithm. Our technique can recognise four kinds of activities which can be used to model part of an individual's current context. We discuss promising experimental results, possible approaches to improve our algorithms, and the impact of this work in modelling user context toward enhanced search and retrieval.**

I. Introduction

The notion of *context* with regard to people using computers, refers to the idea that we can automatically sense characteristics of the environment in which we are and subsequently our computer systems can react in some way to this environment [1], [2]. A good example of context awareness in recent years is the increasing utilisation of location as a source of user context to identify *where* the user is when producing or consuming different media, which is especially important on mobile devices. Indeed, the current generation of smartphones typically include location as a context source when capturing digital photos, which greatly aids retrieval. In addition, we can model the other people *who* are around us via their co-present Bluetooth devices, or *when* it is via timestamps [3]. It is possible to detect some aspects of *what* one is doing using image processing, by analysing the captured media, though this is power-hungry on mobile devices. In this paper we set out to detect some of the *what* aspect of people's current context, we propose classifying accelerometer data to detect the activity an individual is engaged in, which we can then use as input for context-based information retrieval. Specifically we are interested in identifying four different user activities (sitting, walking, driving, lying down). These activities were chosen so as to identify the different user contexts that a mobile device carrier would typically engage in and are activities that could influence how and when information is shown to the user. We capture user activity by analysing accelerometer values from a wearable mobile device, in our case a Microsoft SenseCam (see Figure I), which is a wearable camera worn via a lanyard suspended from the neck, which takes up to 4,000 images per day from a first-person viewpoint. Although the SenseCam is one particular wearable device, our conjecture is that the proposed technique will effectively port to any accelerometer-enabled mobile device.

Fig. 1. SenseCam

The SenseCam, explained in more detail by Hodges *et. al.* [4], captures images and sensory information such as ambient temperature, movement via a tri-axial accelerometer, and ambient lighting. In our work, we focus on the capture of accelerometer data and how this can be used to infer user activity, and following this, how any accelerometer-enabled device can be used. However utilising the SenseCam as a context-gathering device brings with it at least one interesting opportunity for evaluating of the performance of our user activity recognition. A device such as the SenseCam offers a unique opportunity to research activity detection since it captures visually what the wearer is doing and this helps validate the experimental results as it allows for the gathering of an extremely accurate groundtruth for experiments which we believe to be significantly more accurate than an equivalent diary-based record. In the next section we discuss uses of different kinds of context data to support search and retrieval. Following that we discuss our approach to identifying user activities using the wearable device in section II, before presenting our experiments and results in section IV. In section V we outline potential uses of knowing the user activities at both indexing and query time before concluding in section VI.

II. Background

With the increasing ubiquity of mobile devices in our lives, large amounts of multimedia data and also sensor data can be produced easily. How to access such information is a focus of increasing attention, and many researchers are considering the challenges of managing such archives of personal data. [5]. Many techniques are being developed to extract context from all kinds of mobile device data [6], [7], [8] and we focus on one such device, SenseCam, although, as previously mentioned, the proposed technique could translate easily into other accelerometer enabled devices.

The SenseCam is a wearable computing device that is worn around the neck and therefore facing towards the majority of activities the user is engaged in. More importantly for this work, the device is in touch with the person al all times and experiences the majority of the whole-body movements that the person makes. The SenseCam's main function to to visually capture life experiences by taking photos automatically a few times a minute, and it is widely used in the lifelogging community as a data-gathering tool. In addition to having a camera, SenseCam also includes a number of other sensors; the one we are interested in is the tri-axis accelerometer that captures a reading every second. The accelerometer plays an important role in the SenseCam through determining the optimal time to take a picture so as to avoid blurring that would otherwise be prevalent in a moving wearable camera. In our experiments, SenseCam allows us to precisely annotate when various activities occurred, and then build automatic classifiers on this highly accurate groundtruth. This is an important aspect of this work that allows our experiments to take place over an extended period of time and allows the wearer to engage in normal activities while gathering data. Indeed, from our experience, a wearer very quickly forgets that they are wearing SenseCam after putting it on and wearing the device will not impact on the user's daily activities.

There is prior research into the use of accelerometers to identify activities, but in most cases, this research involved the use of several independent accelerometers at various locations on the body [9]. The rationale for using one accelerometer contained in SenseCam is that we envisage a single device (e.g. a mobile phone) being used to gather user activity context data. It is unlikely and unrealistic for a real-world user to wear a number of strategically placed accelerometers to gather context data as part of everyday life. Our conjecture is that we must not expect the user to do anything out of the ordinary in their daily life, and that technologies must adapt to the user's life as opposed to the user adapting to the technology.

Past accelerometer-based experiments have been carried out on datasets of just a couple of hours of activity data. In our experiment we validate our accelerometer activity recognition algorithms over a period of one full week use in a free-living, real-world environment. As shown in Figures II and 3, the circle point is when the photo was taken . Most papers talk about using frequency domain features to recognise activities, but this is not applicable in our experiments. The frequency of our accelerometer is 1 Hz (to facilitate longer battery usage), which is not enough to use frequency data working on activity detection. Past research has highlighted how activity recognition classifiers are not yet sufficiently accurate to use in modelling a user's context [10], but we believe that our results are sufficiently positive so as to be of use for annotating user context.

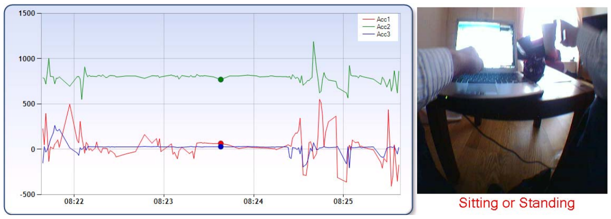


Fig. 2. Graphic of 3-axis accelerations of sitting or standing

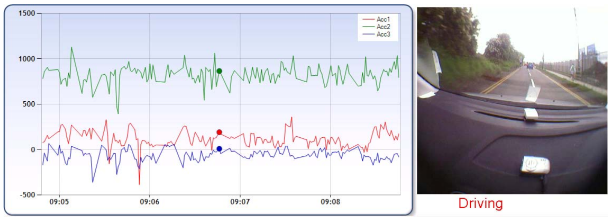


Fig. 3. Graphic of 3-axis accelerations of driving

III. User Activity Context Modelling

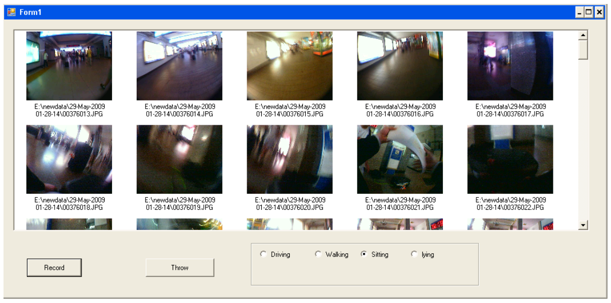
In this section we detail how we can use accelerometer data and machine learning tools to automatically identify user activity. Firstly, however we identify the challenges in classifying a user's current activity using only accelerometer sources. Recall that the four activities we are concerned with are: *Sitting/Standing*,*Walking, Driving and Lying down*.

A. Challenges in accelerometer based activity classification

Classifying activity using accelerometers alone is a challenging task, but even with visual images (as captured with SenseCam), this task is still not straightforward. For example, visually identifying the difference between standing and walking may be impossible. However, when dealing with accelerometer data only, the challenge becomes more acute, for example when the user is waiting for a red light while driving, and thus is not moving (essentially sitting), the acceleration data can have the characteristics of data from sitting. When the user moves over bumps or ramps when driving a car, this may appear like walking. Finally, when the user changes activities between the times of SenseCam images which happens a lot given the frequency with which SenseCam images are taken, issues of boundary definition arise. Our approach to handling these is based on machine learning, whereby we train a Support Vector Machine (SVM) to automatically classify accelerometer features into user activities. This requires the use of a set of underlying features for classification, as now described.

B. Input Features for Activity Recognition

It is impractical to classify the activities only by a single isolated reading of raw accelerometer data taken at the same time as an associated image. To address this we take 10 seconds worth of accelerator readings around every image to extract the relevant features. *Lying down* is the easiest activity to detect among our four activities. Due to gravity, one acceleration of 3-axis is always about 1G, so if the value of this



E:\newdata\29-May-2009 01-28-14\00376013.JPG
E:\newdata\29-May-2009 01-28-14\00376014.JPG
E:\newdata\29-May-2009 01-28-14\00376015.JPG
E:\newdata\29-May-2009 01-28-14\00376016.JPG
E:\newdata\29-May-2009 01-28-14\00376017.JPG
E:\newdata\29-May-2009 01-28-14\00376018.JPG
E:\newdata\29-May-2009 01-28-14\00376019.JPG
E:\newdata\29-May-2009 01-28-14\00376020.JPG
E:\newdata\29-May-2009 01-28-14\00376021.JPG
E:\newdata\29-May-2009 01-28-14\00376022.JPG

○ Driving ○ Walking ● Sitting ○ lying

Fig. 4. Screenshot of our annotation application, which exploits the strength of lifelog images as powerful memory cues to help the annotator in identifying the activity they were engaged in

axis changes to less and another axis increases, our detection algorithm will note the SenseCam's angle with the ground has changed. *Sitting/Standing* is also quite straightforward to detect as when the user is sitting, all the surrounding accelerations exhibit little change. On the other hand *Walking* is a very different activity to classify, as all three accelerations change a lot. An accelerometer has more sensitivity than humans, as when *driving* on the road, even if road is flat people don't detect movement, while an accelerometer still detects minor vibrations.

We use a number of features as input to our activity classifier, and we now describe these:

Raw acceleration data
We can use raw data to judge the posture of the SenseCam. Due to gravity, the value of the accelerometer axis is about 1G. For example, when the user lies down, the value will decrease and another axis's value will increase at the same time.

Standard Deviation
This feature is used to calculate the strength of activities. If the accelerations change rapidly, there is a strong likelihood that the user is walking or driving.

Range
From this feature, we can better distinguish driving from walking. When the user is driving, the *Standard Deviation* may be the same with walking in the same period. However the range of values changing is smaller than for the walking activity.

Because we collect accelerations from a 3-axis accelerometer, a total of 9 features (Raw acceleration data, Standard Deviation and Range for each axis) are used for one reading of acceleration.

C. Activity Classification

We selected the Support Vector Machine (SVM) as a machine learning tool given it's widespread use in classifying accelerometer-based activity [11]. It can be used to classify multi-class data, but in this work we adopt a two-class classification because different classes will be recognized by different features combination.5.

In the process, we classify the training data into two classes (binary classification) for each activity. Following that, we identify the optimal parameters for each of the four activities and then we use the optimal parameters and training data to train the classification model for each activity. Each of the four models are then evaluated using five-fold cross validation.

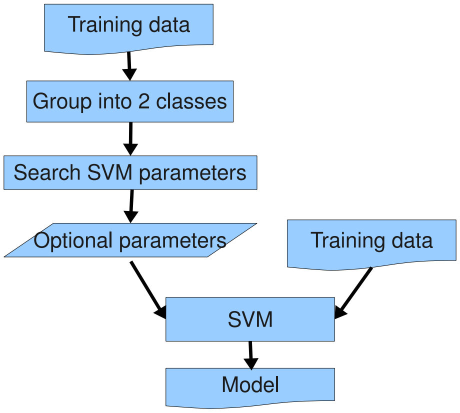


Fig. 5. Process of classifying raw acceleration data into user activities

IV. Experimental Setup

In this section we describe the setup for a test subject who gathered SenseCam data, and then manually annotated ten days of SenseCam images for various activities. This is where the non-accelerometer SenseCam data is important as the visual images (3 per minute) are exactly time-aligned with the accelerometer readings, and therefore we are able to stand over the validity of the user data and are not simply relying on user annotation from memory or diary. Another positive feature is that the user is free to carry on normal daily activities in a free-living environment, therefore our user activities are typical activities that would be carried out on a daily basis anyway. We annotated 17,515 clear photos (activity points) with the four activities *Sitting/Standing, Walking, Driving and Lying down* with the application shown in Figure 4. This application also calculates photo's acceleration attributes with 10 seconds of acceleration data around each photo (over 170,000 accelerometer readings). The manual groundtruth distribution of the 4 kinds of activities is shown in Figure 6. These images, accelerometer readings and groundtruth comprised our test collection for our experiments. We employed five-fold cross validation when training and testing the SVM.

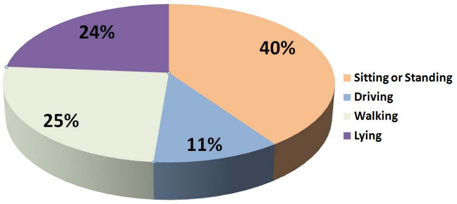


Fig. 6. Number and percentage of pictures manually chosen and annotated.

A. Classification

In our experiments we used LibSVM, an implementation of SVM, and we optimised different parameters to classify

TABLE I
C, γ PARAMETERS SELECTED AFTER 5-FOLD CROSS-VALIDATION TRAINING FOR EACH ACTIVITY

Activity	C	γ
Driving	8	0.0001220703125
Sitting or Standing	8	0.0001220703125
Walking	8	0.0001220703125
Lying	32	0.0001220703125

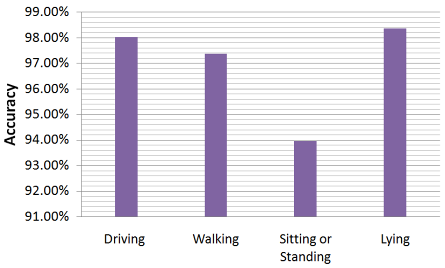


Fig. 7. The accuracy of each activity model (range 90% to 98%).

each of the different activities [12]. We used the RBF kernel with probabilistic output, and optimized parameters C and γ (gamma) in the training phase. The optimised parameters found for each activity are shown in Table I.

B. Results

As mentioned earlier, we trained four models, one for each activity. The accuracy for each activity model is shown in Figure 7.

In Section III-A we discussed the fact that a user typically changes behaviour quite often in everyday life, especially when *standing*, thus explaining why the accuracy of this activity is lower than the other three.

The resulting accuracy of detection of each activity is shown in Table II. As mentioned in Section III-A, 20 *Driving* instances were classified as *Walking* because of uneven road conditions. 144 *Driving* instances were also classified as *Sitting/Standing* because of red lights or stop signs. There are 105 *Walking* instances classified as *Driving*, most likely because of peculiarities in some walking actions for unknown reasons. As behaviours can be changed between periods of photo capture, there are 207 *Sitting/Standing* instances classified as *Walking* and 185 *Walking* instances are classified as *Sitting/Standing*. Given the difficultly in accurately annotating *Lying down* from *Sitting/Standing*,there are 234 *Sitting/Standing* instances classified as *Lying down*. For many of these mis-classifications, a simple post-classification smoothing step would address most of these problems and this is planned for future work.

TABLE II
CONFUSION MATRIX OF EACH ACTIVITY MODEL (ROW IS USER ANNOTATION)

	Driving	Walking	Sitting or Standing	Lying
Driving	**1,647**	20	144	2
Walking	9	**4,066**	185	0
Sitting or Standing	105	207	**6,557**	73
Lying	7	5	234	**3,949**

V. Possibilities & Use Case

The capture of context of the wearer has many uses and applications in information retrieval. As demonstrated by O'Hare *et. al.* we can clearly identify the usefulness of context for indexing of multimedia data, where the semantics of the data will not be as readily available as when indexing text data [7]. Specifically with regard to user activities, one can note that the activity of the user at (or leading up to) photo or video capture time could of course be an important asset in indexing a digital photo or video. Considering e-memories and lifelogging using a device such as SenseCam, then the application of as much context information as possible will greatly aid the search of past digital memories, especially when faced with upwards of 1,000,000 photos in a year which is what a SenseCam can generate.

At query/search time the use of context is also very important, especially when using an interaction limited device such as a mobile phone or a TV. Taking the TV as an example because of its reduced user interaction, activities the user is currently engaged in and how long the user is likely to be able to watch the TV is likely to become important when the TV can access web content and generate personalised playlists for the viewer. When considering mobile devices, the user is faced with the challenge of restricted input modalities and a small screen which does not afford the possibility to engage in complex screen-based manipulation of content. In this case, user context is important and being able to identify the activity the user is engaged in, which would be very useful in that the presentation or the push of data so the use can be tailored to the user's activity and environment. For example, an important news story can be presented to the user in audio format only if the user is driving, but if the user is sitting, video or text presentation could be more suitable.

For one example use case, we present our concept of how capturing user activities are important when dealing with one example usage scenario, that of e-memories.

A. E-memories

Lifelogs or e-memories attempt to capture digitally all aspects of a person's life. This is typically achieved using wearable sensors such as mobile phones, SenseCams, etc. One of the key challenges facing the lifelog research community is that of effectively supporting user search through the lifelog data [13], especially when the user is unlikely to manually annotate the data due to the vast quantities gathered. To effectively use lifelogs and e-memories, we need to better understand what people were doing when the lifelog was captured, so as to provide automatic annotations. To improve search and recall from such lifelogs we want to use a number of context reinstatement techniques to trigger this recall. Both these target motivations require us to capture the *who, what, when, where*, and *why* of our activities. No single source of evidence can successfully provide information on all these facets of activity, but a range of techniques fused together shows promise for providing a solution. Bluetooth can be used to detect other devices in one's vicinity to detect *who* is nearby, GPS can record *where* we are. Images may give an indication of *what* we were doing, but the well addressed problem of the *"semantic gap"* means this solution is still some time away from maturity [14]. Therefore the role of accelerometers in quickly and accurately identifying the *"what"* aspect of our context may well provide a shorter-term solution to better support our access to e-memories and lifelogs.

Consider the scenario of a typical afternoon in the life of John as illustrated in Figure 8. These high level activities of *"finishing work in the lab", "at the bus stop", etc.* can be naturally broken down into *sitting, walking, driving, etc.* which our accelerometer based processing can detect. The value in understanding a user's contextual situation (e.g. sitting, walking) is helpful from an information retrieval perspective, as it can be used by John e.g. *"find me the occasions when I was at the Grand Hotel, after walking there"*. Also in real-time John will have the facility to be presented with past (related) e-memories of other walking activities around the Grand Hotel, e.g. when he went for a walk with his friend Alice in a nearby park.

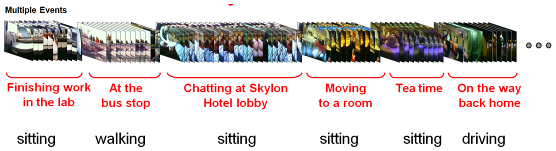


Fig. 8. Example of typical activities a user is involved in.

VI. Conclusions & Future work

In this paper we have illustrated how we can use a wearable accelerometer to identify the activities of a wearer to a very high accuracy. We have employed a SenseCam for this work, but equally any accelerometer-enabled device (e.g. a mobile phone) could be successfully employed. We have chosen for this work to identify four activities, but the identification of additional activities can be explored and requires only the training of additional classifiers. Our belief is that we can train additional classifiers and in the majority of cases, that we can maintain equivalent performance to the classifiers already described.

There are a number of future research opportunities that we are addressing:

- New activities to be identified from the acceleration data;
- Adopting smoothing algorithms to improve accuracy. For example driving can be misclassified on the micro level because of the stop-start nature of driving;
- Investigating high-frequency accelerometers that can give more information about body movements, while considering possible battery lifespan trade-offs.

When considering new activities, in addition to the four activities just described, we are looking at recognising *flying* from *sitting/standing*, classifying *driving* into *train* and *car*, and also to identify *running* vs. *walking*. These more fine-grained activities will allow for a better understanding of a user's current context, which in future will better assist their information needs at any given time.

Acknowledgements

This material is based upon work supported by Science Foundation Ireland under Grant No. 07/CE/I1147.

References

[1] H. W. Gellersen, A. Schmidt, and M. Beigl, "Multi-sensor context-awareness in mobile devices and smart artifacts," *Mob. Netw. Appl.*, vol. 7, no. 5, pp. 341–351, 2002.

[2] L. Barnard, J. S. Yi, J. A. Jacko, and A. Sears, "Capturing the effects of context on human performance in mobile computing systems," *Personal Ubiquitous Comput.*, vol. 11, no. 2, pp. 81–96, 2007. [Online]. Available: http://portal.acm.org/citation.cfm?id=1229065

[3] A. Sorvari, J. Jalkanen, R. Jokela, A. Black, K. Koli, M. Moberg, and T. Keinonen, "Usability issues in utilizing context metadata in content management of mobile devices," in *Proceedings of the third Nordic conference on Human-computer interaction*. Tampere, Finland: ACM, 2004, pp. 357–363.

[4] S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur, and K. Wood, "Sensecam: A retrospective memory aid," in *UbiComp: 8th International Conference on Ubiquitous Computing*, ser. LNCS, vol. 4602. Berlin, Heidelberg: Springer, 2006, pp. 177–193.

[5] K. Church, B. Smyth, P. Cotter, and K. Bradley, "Mobile information access: A study of emerging search behavior on the mobile internet," *ACM Trans. Web*, vol. 1, no. 1, p. 4, 2007. [Online]. Available: http://portal.acm.org/citation.cfm?id=1232726

[6] Y. H. Yang, P. T. Wu, C. W. Lee, K. H. Lin, W. H. Hsu, and H. H. Chen, "ContextSeer: context search and recommendation at query time for shared consumer photos," in *Proceeding of the 16th ACM international conference on Multimedia*. Vancouver, British Columbia, Canada: ACM, 2008, pp. 199–208. [Online]. Available: http://portal.acm.org/citation.cfm?id=1459387

[7] N. O'Hare, C. Gurrin, G. J. F. Jones, H. Lee, N. E. O'Connor, and A. F. Smeaton, "Using text search for personal photo collections with the MediAssist system," in *Proceedings of the 2007 ACM symposium on Applied computing*. Seoul, Korea: ACM, 2007, pp. 880–881. [Online]. Available: http://portal.acm.org/citation.cfm?id=1244195

[8] L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury, "How flickr helps us make sense of the world: context and content in community-contributed media collections," in *Proceedings of the 15th international conference on Multimedia*. Augsburg, Germany: ACM, 2007, pp. 631–640. [Online]. Available: http://portal.acm.org/citation.cfm?id=1291384

[9] *Activity Recognition from User-Annotated Acceleration Data*, April 2004.

[10] A. R. Doherty and A. F. Smeaton, "Automatically augmenting lifelog events using pervasively generated content from millions of people," *Sensors*, vol. 10, no. 3, pp. 1423–1446, February 2010.

[11] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," *American Association for Artificial Intelligence*, 2005. [Online]. Available: http://paul.rutgers.edu/˜nravi/accelerometer.pdf

[12] "LIBSVM – a library for support vector machines," http://www.csie.ntu.edu.tw/˜cjlin/libsvm/. [Online]. Available: http://www.csie.ntu.edu.tw/ cjlin/libsvm/

[13] K. O'Hara, M. Tuffield, and N. Shadbolt, "Lifelogging: Issues of identity and privacy with memories for life," in *The First International Workshop on Identity and the Information Society*. Berlin / Heidelberg, Germany: Springer, 2008, pp. 1–31.

[14] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1349–1380, Dec 2000.