

USING BLUETOOTH & GPS METADATA TO MEASURE EVENT SIMILARITY IN SENSECAM IMAGES

DARAGH BYRNE, BARRY LAVELLE, AIDEN R. DOHERTY,
GARETH J.F. JONES, ALAN F. SMEATON

*Centre for Digital Video Processing (CDVP) & Adaptive Information Cluster(AIC),
Dublin City University, Dublin 9, Ireland*

The Microsoft SenseCam is a small multi-sensor camera worn around the user's neck. It was designed primarily for lifelog recording. At present, the SenseCam passively records up to 3,000 images per day as well as logging data from several on-board sensors. The sheer volume of image and sensor data captured by the SenseCam creates a number of challenges in the areas of segmenting whole day recordings into events, and searching for events. In this paper, we use content and contextual information to help aid in automatic event segmentation of a user's SenseCam images. We also propose and evaluate a number of novel techniques using Bluetooth and GPS context data to accurately locate and retrieve similar events within a user's lifelog photoset.

1. Introduction

Lifelogging is a term used to describe the notion of a person digitally capturing his or her life experiences. There can be many different forms of capture including a record of one's e-mail messages, web pages explored, music listened to, personal photographs and personal video. Lifelogging is a growing phenomenon with many people interested in recording their life's activities for posterity, for calendar, medical and diary applications and for subsequent nostalgic browsing. This paper explores an aspect of reviewing one's personal photographs in the lifelogging domain.

To aid the capture of digital images representing a user's lifelog, we use a device developed by Microsoft Research in Cambridge, U.K., known as the SenseCam [1]. The SenseCam is a small wearable device which incorporates a digital camera and multiple sensors detecting changes in light levels, motion, and ambient temperature. There is also a passive infra red sensor to detect the presence of an individual. Sensor data is captured approximately every 2 seconds and stored on-board. Based on these readings it is determined when an image should be captured.



Figure 1 SenseCam worn by user

At present, the SenseCam passively captures up to 3,000 photographs per day, thus building up an extensive lifelog of images for an individual. There is a substantial research challenge in managing this sizeable collection of over 20,000 images per week which equates to approximately 1 million images captured per year. Over a lifetime of wearing this passively capturing camera, a user could expect to have a collection

of over 50 million images. In such circumstances, it is unreasonable to expect a user to manually search through images encountered from their entire lifelog without the help of automated segmentation of images into events and retrieval of those events. To address this issue, it is necessary to segment each day's images into a series of distinct activities, as illustrated in figure 2. Once those activities are identified it will thereafter be necessary to retrieve similar events to a reference event. Given the ubiquity of Bluetooth technology [2], in this paper we investigate various approaches to finding similar activities based on passively recorded Bluetooth and GPS contextual information.

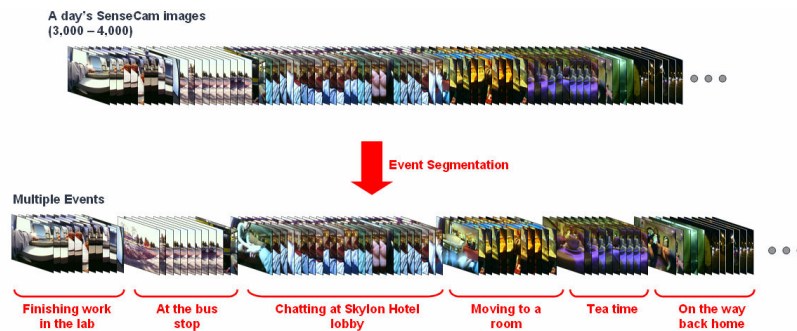


Figure 2 Segmentation of images into events

Section 2 will describe related research to our work in terms of retrieving similar events. In section 3 we describe how we segment our SenseCam images into distinct events. In section 4 we propose various approaches utilising Bluetooth and GPS information to retrieve similar events. Sections 5 and 6 describe the experimentation and evaluation of our contextual retrieval techniques.

2. Related Work

Previous work by Ellis and Lee [3] performs clustering on detected events to infer user activities, however they work in the audio domain only. Wang *et. al.* [4] use the event to query by example to find other similar events based on visual and audio features, as do Lin and Hauptmann [15]. Several studies have investigated the role wearable sensors play in the automatic classification of human activities [5][12][13]. In our approach, we investigate the retrieval of similar events based on contextual information alone, namely using Bluetooth devices and GPS co-ordinates logged during the event. Both of these technologies are widespread and readily available in mobile devices [2,6]. Hence they provide pragmatic sources for our experimentation, as suggested by Bristow *et. al.* [14] in their study to define the different types of contextual information

that are important in wearable computing. We are not aware of similar work using Bluetooth for this purpose.

3. Segmenting images into distinct events

Based on previous work [8] we segment our images into distinct events or activities by making use of a combination of three different sources of information: low-level image features (content), light level sensor (context) and accelerometer/motion sensor (context).

3.1 Content processing

To segment a day's worth of images into distinct events based on the image content, we make use of the aceToolbox [7] to calculate 5 low-level MPEG-7 feature descriptors. Adjacent blocks of images are compared against each other and where there is a sufficiently large difference between two adjacent blocks, a potential event change is logged. For example, if the wearer is at breakfast and then walks out to get the bus to work, there will be a significant change in the visual properties of the captured images, which may trigger an event change.

3.2 Context processing

Two different contextual data sources are used, namely the onboard light sensor and motion sensor. For example if the wearer is sitting down at work in front of their PC and then decides to walk to lunch, there will be a significant change in motion activity, which may trigger an event change. Similarly as the wearer moves from indoors to outdoors to walk to lunch, there will be a significant change in the level of lighting sensed. Each image is associated with light and motion values, and as with our content processing we search for distinct changes in sensor values.

4. Event Retrieval

One of the issues faced in retrieving similar events to a reference event is the question of how to effectively utilise context data. In our experiments, one of the authors wore a SenseCam, logged his location via a handheld GPS device and recorded friendly names and MAC addresses of Bluetooth devices in his vicinity over a 24 day period. The gathered data was analysed and segmented into discrete events and we evaluated a number of techniques, which we discuss below, to help determine how similar any two events are.

4.1 GPS

GPS co-ordinates were logged by the user in conjunction with the SenseCam and Bluetooth context data. GPS offers us a means of determining the location at which an event occurs. For each event we calculate the distance in kilometres

(Km) between it and every other event in the set and use this to determine how similar events are to each other.

4.2 Bluetooth device presence

Each event contains a set of Bluetooth devices which were present. To provide us with a basic similarity score, the device set for one event is compared to that of another using the *Jaccard co-efficient* [9], a recognised means for determining similarity. The intersection of events being compared represents those devices which were co-present for both events.

$$J(A, B) = |A \cap B| / |A \cup B|$$

The results yielded by this approach provide a similarity score in the range [0,1], with scores closer to 1 indicating a high level of similarity between sets. As will be described we can combine the results of this with a number of other factors including the duration for which a device was present during an event and the familiarity of the Bluetooth device for a particular user.

4.3 Bluetooth duration

We calculate the duration of each device present in an event to up-weight those devices that were in attendance the longest. Our belief being that the greater the proportion of an event that a device is present for, the more significant the device (or its owner) is to that event. The following formula is used to weight the similarity based on duration.

$$DurationWeight = \frac{|X \cap Y| - \sum_{i=0}^{|X \cap Y|} DiffDur(X_i, Y_i)}{|X \cup Y|}$$

$$DiffDur = |Duration(X_i) - Duration(Y_i)|$$

$X = \text{Event 1}$, $Y = \text{Event 2}$, $i = \text{devices present in both events}$

4.4 Device presence weighted by familiarity

In previous work [10], we address the concept of assigning familiarity scores to Bluetooth devices, the familiarity score being a measure of device presence relative to the other devices encountered within the set. Including familiarity as a weight promotes those events in which familiar devices were encountered. We believe this may be useful for finding similar events in which familiar people are known to the user and are likely to be present in.

4.5 Devices weighted by duration and familiarity

Another approach combined work in 4.3 and 4.4 (above) to examine the effects

of combining duration with familiarity. Combining these attributes gives us a way of detecting how well the user knew those present in a particular event and also what duration they were present for. In theory, if people who are well known by the user appear in an event then the similarity rating should up-weight comparative events in which the same people occur, for a similar period of time.

4.6 Devices weighted by inverse familiarity

This was designed to up-weight strangers and outliers in events, in a similar manner to TF-IDF [9] in information retrieval. This approach gives precedence to those events in which co-present devices with low familiarity scores occur. This allows us to detect similar events based on non-familiar users. Examples of where this may be relevant would include times where the user encountered the same set of relatively unfamiliar users e.g. meetings with an infrequent acquaintance.

5. Experimental Setup

To aid in the collection of Bluetooth context data, we employed a Java Mobile Edition (Java ME) application used in previous work [10] to log nearby Bluetooth devices and also capture a time-stamp for each time a device was encountered. The logger was run on a mobile phone, in conjunction with the SenseCam and a hand-held GPS device. The time-stamp of digital images captured was synchronized with that of the Bluetooth logger and GPS device. To evaluate the proposed techniques from the previous section, 10 random events were firstly selected from the user's dataset of approximately 25,000 images. A reference event was selected and remaining events were then ranked based on their similarity to this. Users were provided with a summary of the top 10 highest ranking events. Three key frames were automatically selected for each event and displayed via a web page. A judgement was then made by the user to indicate whether the displayed events were similar to the reference event or not. While this approach does not provide recall values, it does afford precision values which indicate what approach is likely to perform best for certain event types.

6. Results

By associating a particular device to a user we can infer the presence of specific individuals in an event. Therefore Bluetooth context enables person based retrieval of similar events. This is a possible advantage over traditional content based retrieval. We found contextual information to be significantly faster than low-level feature analysis as it offers near real-time processing & indexing. Table 1 below provides us with a breakdown of the results achieved in our experimentation. By classifying events based on motion, we can see that performance is dependent on the type of event and context data used.

Table 1. Precision@10 for randomly selected events

Event	Motion	GPS	BT Activity	BT Duration	Familiarity	Familiarity & Duration	Inverse Familiarity	BT Activity & GPS
1	High	1.0	0.2	0.2	0.2	0.2	0.4	0.4
2	High	0.9	0.2	0.2	0.1	0.1	0.7	0.6
3	High	0.9	0.9	0.9	0.9	0.8	0.8	0.8
4	Low	0.1	0.4	0.4	0.2	0.0	0.3	0.2
5	Low	0.1	0.1	0.2	0.1	0.0	0.4	0.3
6	Low	0.1	0.9	1.0	1.0	0.9	0.7	0.8
7	None	0.5	0.8	0.8	0.9	0.8	0.7	1.0
8	None	0.1	0.2	0.2	0.2	0.1	0.3	0.2
9	None	0.7	0.7	0.7	0.8	0.8	0.7	0.8
10	None	0.0	0.1	0.1	0.1	0.0	0.1	0.0
Avg.		.44	.45	.47	.45	.37	.51	.51

From the table above we note that GPS performs particularly well for cases of high motion. A high motion event is one where a large number of different GPS coordinates have been encountered. Traveling to and from work, for example, would be considered a high motion event whereas time spent working at a desk, would be a no motion activity (having only one GPS co-ordinate). Conversely, GPS yielded mixed results for low and non-motion events. This was due to the large majority of events occurring in a relatively small geographical area i.e. the user’s place of work. This made it difficult to accurately retrieve similar events. As such, GPS doesn’t appear to provide useful context data for similarity matching where similar events occur in close proximity to one another.

We found Bluetooth information performed consistently well irrespective of weighting approach used. On further examination the results appear to be quite polarised between high and low levels of precision. We can account for high motion events being dissimilar as it is highly unlikely to encounter a large enough set of similar devices occurring across high motion events. Given that it takes approximately 10 seconds to complete a Bluetooth device discovery, it is possible for devices to move out of range and go undetected in these cases.

7. Conclusions & future work

We have demonstrated that context information provides a useful alternative for retrieving similar lifelog events. We believe that the true benefit of our work will be realised when using low-level content in combination with contextual sources. Our results indicate that the Inverse Familiarity weighting and the Bluetooth combined with GPS approach prove most promising for retrieving similar events. In our future work, we plan to use the rate of change in Bluetooth activity to enrich our current approaches of segmenting images into events.

Acknowledgements

We are grateful to the AceMedia project and would like to thank the Irish Research Council for Science, Engineering and Technology; and Science Foundation Ireland under grant number 03/In.3/1361 for support.

References

1. 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 *Proc. UbiComp 8th International Conference on Ubiquitous Computing*, (2006)
2. <http://www.bluetooth.com/Bluetooth/SIG/Billion.htm> (November 2006)
3. D. P. Ellis and K. Lee, "Accessing minimal-impact personal audio archives," *IEEE Multimedia* 13, (2006)
4. Z. Wang, M. Hoffman, P. Cook, and K. Li, "Vferret: Content-based similarity search tool for continuous archived video," in *CARPE Third ACM workshop on Capture, Archival and Retrieval of Personal Experiences*, (2006)
5. Parkka, J. Ermes, M. Korpipaa, P. Mantyjarvi, J. Peltola, J. Korhonen, I., "Activity classification using realistic data from wearable sensors", *IEEE Transactions on Information Technology In Biomedicine*, Vol. 10, No. 1, (January 2006)
6. <http://www.fcc.gov/911/enhanced/> (May 2007)
7. "The AceMedia project", <http://www.acemedia.org>, (April 2007)
8. A. R. Doherty, A. F. Smeaton, K. Lee, and D. P. Ellis, "Multimodal segmentation of lifelog data," in *Proc. RIAO 2007, Pittsburgh*, (2007) (to appear)
9. C. J. Van Rijsbergen, "Information Retrieval", *Butterworth-Heinemann*, ISBN: 0408709294, (1979)
10. B. Lavelle, D. Byrne, G. J. F. Jones, C. Gurrin, A. F. Smeaton, "Receive Data From?...A mechanism for determining familiarity amongst Bluetooth devices", submitted to *Mobile HCI 2007, Singapore* (2007)
11. R. Baeza-Yates, B. Ribeiro-Neto, "Modern Information Retrieval", *Addison-Wesley*, Wokingham, UK, (1999)
12. M. Philipose, K. P. Fishkin, M. Perkowitz, D. Patterson, and D. Hahnel, "The probabilistic activity toolkit: Towards enabling activity aware computer interfaces", tech. rep., Intel Research Seattle Technical Memo, (2003).
13. W. Pentney, A.-M. Popescu, S. Wang, H. A. Kautz, and M. Philipose, "Sensor-based understanding of daily life via large-scale use of common sense." in *AAAI*, (2006).
14. H. W. Bristow, C. Baber, J. Cross, J. F. Knight, and S. I. Woolley, "Defining and evaluating context for wearable computing", *International Journal of Human Computer Studies* 60, (2004).
15. W. Lin and A. Hauptmann, "Structuring continuous video recordings of everyday life using time-constrained clustering," in *Multimedia Content Analysis, Management, and Retrieval SPIE-IST Electronic Imaging*, (2006).