

Combining Face Detection and Novelty to Identify Important Events in a Visual Lifelog

Aiden R. Doherty and Alan F. Smeaton
Centre for Digital Video Processing and Adaptive Information Cluster
Dublin City University, Glasnevin, Dublin 9, Ireland.
adoherty@computing.dcu.ie

Abstract

The SenseCam is a passively capturing wearable camera, worn around the neck and takes an average of almost 2,000 images per day, which equates to over 650,000 images per year. It is used to create a personal lifelog or visual recording of the wearer's life and generates information which can be helpful as a human memory aid. For such a large amount of visual information to be of any use, it is accepted that it should be structured into "events", of which there are about 8,000 in a wearer's average year. In automatically segmenting SenseCam images into events, it is desirable to automatically emphasise more important events and decrease the emphasis on mundane/routine events. This paper introduces the concept of novelty to help determine the importance of events in a lifelog. By combining novelty with face-to-face conversation detection, our system improves on previous approaches. In our experiments we use a large set of lifelog images, a total of 288,479 images collected by 6 users over a time period of one month each.

It captures approximately 2,000 images on an average day creating a sizable collection of images even within a short period of time, i.e. over 650,000 images per year (or over 40 million in a lifetime). In our previous work [5] we introduced an approach to manage such a large visual dataset by automatically splitting these collections into manageable segments by identifying the boundaries between different activities in the wearer's day, e.g. having breakfast, working in front of a computer, attending a game of football, etc. (Figure 2)

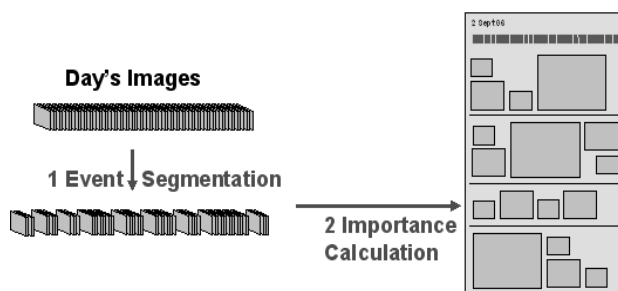


Figure 2. Determining Event Importance.

1. Introduction

The SenseCam (Figure 1), developed by Microsoft Research in Cambridge, UK, is a small wearable device which incorporates a digital camera and onboard sensors (motion, ambient temperature, light level, and passive infrared to detect presence of people).



Figure 1. SenseCam worn around neck.

On average a user will have approximately 22 distinct events in a day (i.e. around 8,000 events per year), but it is quite likely that many of these events may be routine or repeated events that are relatively unimportant to the user. However there will also be certain events that may be of great significance to the user, e.g. lunch with a loved one, an awards ceremony, attending a big football match, etc., because they are more unique or novel. Thus it is desirable to automatically emphasise the events that may be more unique to the user than the routine or mundane events, such as is abstracted in Figure 2. This paper extends the state of the art of determining the importance of lifelog events/activities through the use of the uniqueness or novelty of all events in a user's lifelog, coupled with automatic detection of faces and of conversation.

2. Related Work

The field of lifelogging is a relatively new but rapidly expanding area of research [3, 8]. Hodges *et. al.* detail the potential memory benefits of a personal visual lifelog such as that generated by the SenseCam [7]. Our approach to visual lifelogging is to capture images throughout the wearer’s day rather than just recording special events or happenings. This generates a true, full lifelog and in previous work we segmented a rather large lifelog of data into manageable shots/activities/events [4].

To our knowledge no groups have captured data for the duration of an entire day, nor from a range of users. Other published work has dealt with data collected from just one user (capturing a wearer’s images for less than 6 hours a day) [8, 9]. For this paper, 6 users captured SenseCam image data for an average of 10 hours per day over a one month period. This provides a more thorough lifestyle representation over a number of users.

Determining of the importance of events in the domain of lifelogging was first introduced by Tancharoen and Aizawa [9]. In their work they determine conversational scenes as being very important events. To achieve this automatically, they used automatic face detection to determine events containing face-to-face conversation. While the use of audio recording may seem obvious to detect conversational scenes, during the collection of data in previous work [5], we found that many people were exceptionally uncomfortable in having the audio content of their conversations recorded.

Aizawa *et. al.* describe their approach to summarising wearable video in which scene importance is calculated based on the alpha wave signal from a wearer’s brain sensor [2]. However this requires the user to wear a brain wave sensor on their head which can affect how the user interacts socially with others. The SenseCam is small and light and from experience of wearing the device, after a short period of time, it becomes virtually unnoticed to the user or to others.

In this paper, the value in detecting the novelty [6] of lifelogging events will be explored, based on the hypothesis that those events that are novel (e.g. rarely occurring events such as going to a rock concert) will be more important to a user than those events that occur frequently (e.g. having breakfast in the morning), because they are more unique, at least more unique visually.

3. Event Importance Approaches

In our lifelog work, each image is represented by MPEG-7 descriptor values plus a range of sensor readings from the on-board SenseCam sensors taken at the same time as the

image. The task we investigate is segmenting these into distinct events [5].

In this paper we investigate three techniques for calculating event importance:

- Our method of novelty calculation for each event
- The conversation detection approach based on face detection [9]
- Importance determination based on a combination of conversation detection and novelty

3.1. Event Importance Based on Novelty

In order to detect novelty we need to compare each event against each other, even naively, and to determine which event(s) are most different to the rest. This requires us to compare pairs of events against each other, and we do this by representing each event by the average value of all the SenseCam images making up that event. These event representative features are then compared against each other using a combination of MPEG-7 visual features, accelerometer, light, passive infrared, and temperature sources. In a post-processing step, events containing a large number of uniformly dark images are assigned a novelty score of zero.

Five separate novelty detection approaches which vary the “window” for novelty detection were evaluated in this paper as follows:

Previous 7 Days To determine the importance of a given event in a day, we investigate how dissimilar each event is to all the other events of that day and of the previous 7 days. The premise of this approach is to take into account the range of activities occurring during the week, which should contain a number of routine/mundane events (breakfast, being at work, etc.) as well as identifying the outlier events such as a barbecue the wearer may have had on a Saturday afternoon.

Previous 3 Days This approach has half of the processing load of the *previous 7 days* approach and may highlight more strongly an important social event that occurred at home during the weekend, as it would be different to the previous 3 days of work. By considering the previous 7 days such an event may have not looked so unique as it would be quite similar to the events from the previous weekend.

Previous 14 Days This approach is more computationally expensive, but a possible advantage will be to decrease the novelty of regular events after a brief break from an individual’s routine. Consider an employee who works one week, then goes on holiday for a week, and then back to work the following Monday with a social

gathering that evening. By considering the previous 14 days it can be seen that in the week prior to being on holiday, the employee spent a considerable amount of time in work, thus the social event of that Monday evening would then score as being more highly novel than the day at work.

Events Within ± 2 Hours From Previous 7 Days This is a more sophisticated approach that considers only events from the previous 7 days that are within ± 2 hours of the event we are trying to compute the novelty of. The premise for this is to highlight, for instance, a family meal out in a nice restaurant at the evening time rather than the normal evening meal at home.

Previous 21 Days Constrained to Same Week Day Here we only consider events occurring on the same day of the week over the past 3 weeks e.g. A doctor may attend a meeting every Thursday, and when data from only the previous week is considered this may appear unique, however when data is taken from only the previous 3 Thursdays, this may rightly appear as a routine event.

3.2. Calculating Event Importance Based on Detecting Conversations

Conversation detection is based on the presence of faces in SenseCam images, on the premise that when talking to someone we are directly facing them, thus this can be determined via the automatic detection of faces from SenseCam images.

We used the Intel OpenCV [1] face detection implementation. A number of face detection trained classifier models are provided with this library and other parameters can also be tuned such as the image scaling factor and the window size that a face should be present in. As SenseCam images are of low quality (generally 25kB each) and taken using a fisheye lens, we investigate which of these models is most suitable for lifelog images.

The importance value for each event is determined by the percentage of event images that have faces detected. Higher scores indicate longer face-to-face conversations, and thus more importance should be attached to such events.

3.3. Calculating Event Importance by Combining Conversational Detection and Novelty

Considering the approach just mentioned, it is highly likely that there will still be many events which will have a face count score of zero yet which are novel! Therefore it is important to take these into consideration and rate the

User	Total Num Images	Num Events	Num Days	Daily Duration
1	79,595	1071	35	14h 54m
2	54,823	661	28	12h 29m
3	42,557	405	20	12h 02m

Table 1. User Statistics

uniqueness of these events to determine which are least interesting. A result of this rationale is that another approach will be investigated whereby the importance of an event is ranked firstly by the concentration of images with faces present, and secondly ranked by the image-based novelty score of that event.

4. Experimental Setup

Six users (in their early twenties to mid thirties) each wore a SenseCam over a one month period, collecting 288,479 images, representing 3,445 events in total. However as novelty calculation is dependent on considering data from previous days, our experiments on novelty detection were limited to periods of time when a user has been wearing the camera for a number of consecutive days. After initial trials, it was decided to only consider days where a user had been collecting data for at least 14 consecutive days without break. This resulted in a dataset of 176,975 images collected from 3 users, which still represents a large dataset of lifelog images.

Table 1 provides a detailed breakdown of the 176,975 images captured by the 3 users. To determine the optimal face detection parameters, the performance of various models was measured against a groundtruth of 1,758 SenseCam images from one day of data from user number 2 in Table 1.

Given the unique and highly personal nature of lifelogging images, the owners of the various SenseCam data collections were the sole judges of the various experiments on their own datasets.

Given the subjective nature of rating how important an event is in relation to other events, it is very difficult to rank the importance of all events within a day, and to do this for each and every day would present a large annotation burden on users. Therefore given that it is of much interest to generally determine the *most* interesting events in a given day, in addition to determining the *most* mundane/routine events from a day, a decision was made to present the two most important and two least important (as determined by the approach under investigation) events to the user and ask them to give a Likert judgement on how much they agree with the proposed events as a summarisation of that day.

Figure 3 illustrates the application built for our users to judge the results from the various novelty detection experiments approaches on their own data. The top two boxes



Figure 3. Groundtruthing interface.

display a selection of images from the two most important events, while the two blue boxes at the bottom of the screen display a selection of the images from the two least important events in the day (as rated by our proposed system). All the images in the middle are keyframe images from each event in that day. The two most and least important events are colour-coded among all those event keyframes, and thus their temporal position relative to all the other events in the day is highlighted.

In total 949 judgments were made by all 6 users, with 664 judgments from 3 users being analysed in the next section.

5. Results

In the first experiments we investigate the best novelty detection approach and thereafter the optimal face detection approach is determined. Finally three systems are compared: 1) The best novelty approach, 2) The best face/conversation detection approach, and 3) Face detection and novelty combined.

5.1. Determining Best Novelty Approach

Five different “windows” of past events against which the novelty of a given event was calculated, were investigated, as detailed in Section 3.1. Figure 4 summarises the performance of these, illustrating the variation in performance for different users for each approach. The middle dots joined up by the thin line indicates the average value across users for each approach. Figure 4 poses 3 questions:

- The optimal number of previous days data to consider in the window of novelty detection ?

- Is it beneficial to look only at events that occur at a similar time of day ?
- Is it advantageous to only consider events from the same day of the week ?

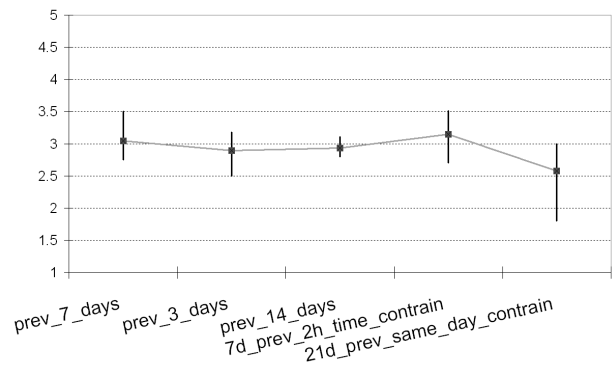


Figure 4. Summary of User Likert Judgements on Novelty Approaches

Number of previous days to consider?

Taking the overall average of each approach, considering the previous 7 days of data is most beneficial with an overall average Likert score of 3.05 (7 days) vs. 2.94 (14 days) vs. 2.89 (3 days). However on close examination of Figure 4 we recommend that the optimal approach is to consider the previous 14 days of data. The extra processing load to compute this approach is not a significant drawback (processing for all approaches being practically instant). By considering events from a greater number of previous days (i.e. the previous two weeks) a more accurate model of a user’s lifestyle can be considered, while allowing for changes in a user’s routine too.

Considering event novelty with respect to time of day ?

A comparison is made here between the *prev-7-days* approach which works marginally better than the *7-days-prev-2h-time-constrain* approach in Figure 4. An advantage of the time-constrained approach is that it considers the novelty of an event not only based on how different it is to all the other events, but also by considering the time of day when that event took place, e.g. a meeting late in the evening may be very important and unique to the individual, whereas it may not be so important or unique if it occurred during regular working hours. In conclusion we recommend that it is advantageous to calculate the novelty of a given event by only considering comparisons to events

from previous days that occur within ± 2 hours of the event in question.

Only considering data from same weekday ?

Inspecting Figure 4 it is evident that the approach *21-day-prev-same-day-constrain* is the least effective approach. A possible explanation for this is that only 3 days of previous data (e.g. the previous 3 Wednesdays/Mondays/etc.) are considered, and this may mean that an insufficient body of data is accumulated to accurately calculate the novelty of an event. In conclusion we recommend that it is not advantageous to calculate the novelty of an event against events occurring on the same given week-day only.

In summary, based on the evidence presented above, to consider the novelty of a given event it should be calculated how dissimilar this event is to all other events that have occurred within ± 2 hours of this event within the previous 14 days. Events are ranked by dissimilarity value, with the most dissimilar being regarded as the most novel/important or unique event of that day.

5.2. Determining the Best Face Detection Method

As is illustrated in Figure 5, there is a trade-off between precision (dashed line) and recall (dotted line), with approaches sorted in descending order in terms of precision. While it may be intuitive to select the approach with the largest F score (continuous thick black line), we decided to use an approach that has a very high precision score. In terms of calculating the importance of events through determining the number of images with faces present (relative to the duration of that event), many events are likely to have a score of zero, therefore it is not so important to distinguish between less important events (all the zero-scoring events), but to identify those few events that are important. As a result, a greater emphasis should be placed on the accuracy/precision of actual faces detected by the face detection system. The approach with the second highest precision score (*haarcascade-frontalface-alt*, scaling factor of 1.1, 3 neighbours, and window size of 30 pixels) is selected as its precision value is very close to the best precision score (0.6336 vs 0.6577, 3.8% worse), while its recall value is better than the recall value of the approach with the best precision score (0.2872 vs 0.2526, 12.0% better).

5.3. Best Overall Event Importance Approach

As evidenced in Figure 6 the face detection approach works considerably better on average than the best proposed

novelty approach. However even more interesting is the fact that when these approaches are combined together we get an improvement on the current state of the art in this task. The advantage of this approach over the *face detection only* approach is that the novelty approach is very good at detecting the routine/mundane events, while the face detection is very good at detecting the most interesting events of a given day. In fact the *face detection + novelty* approach performs at least as well as all the other approaches on 66 of the 83 days from all three users, which is almost 80% of the time. Overall, on average, this approach has a Likert score 4% higher (3.89 vs 3.75) than the state of the art (face detection only).

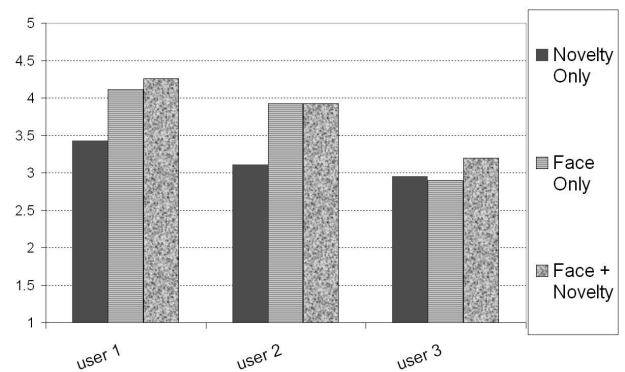


Figure 6. Avg. Likert Score for Main Methods.

It is interesting to note that the introduction of face detection did not have as pronounced an impact on user 3 as it had on users 1 and 2. User 3 was very busy at work abroad during the month he was capturing data, so he did not have as many conversational events during this hectic time. Additionally due to the nature of the light level in certain areas of his office, many images were falsely detected as having the presence of a face, as shown in Figure 7.

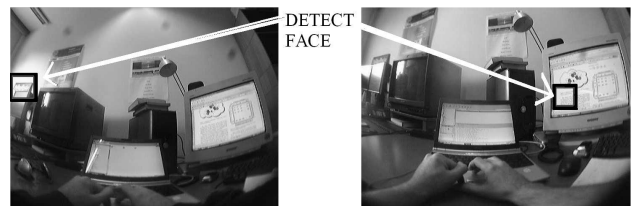


Figure 7. False Positive Face Detection.

6. Conclusions

A number of approaches to determine the relative importance of events from a visual lifelog were investigated on a

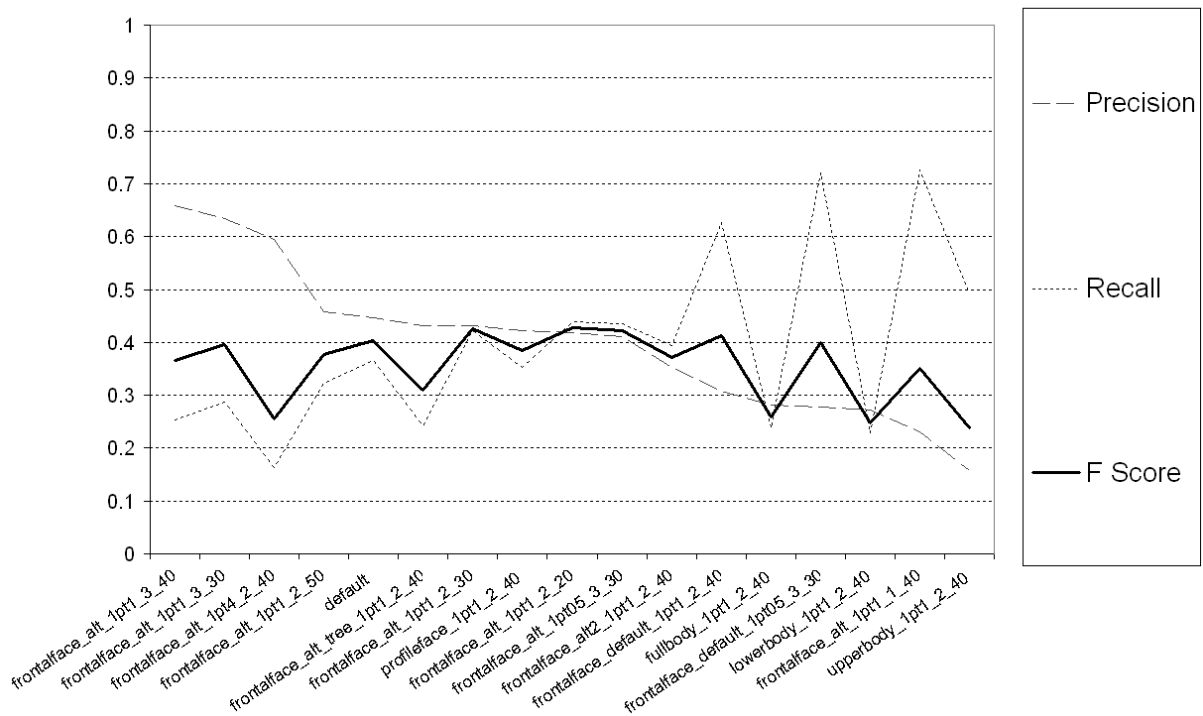


Figure 5. Summary Face Detection Models

dataset of 176,975 images from 3 users. The idea of determining the novelty or visual uniqueness of lifelog events has been introduced into this domain for the first time here. The previous state of the art was based on determining events which have face-to-face conversations as the most important events. However by integrating this with the concept of visual uniqueness or novelty we have developed an approach that works *at least* as well as the previous state of the art 80% of the time and performs 4% better than the previous state of the art overall.

Acknowledgements: The authors wish to acknowledge the support of: Microsoft Research (grant number 2007-056); the AceMedia project; the Irish Research Council for Science, Engineering and Technology; and Science Foundation Ireland (grant number 03/IN.3/I361).

References

[1] Intel opencv. <http://www.intel.com/technology/computing/opencv/index.htm>, Last checked 17 December 2007 2007.

[2] K. Aizawa, K.-I. Ishijima, and M. Shiina. Summarizing wearable video. In *Image Processing*, 2001.

[3] G. Bell and J. Gemmell. A digital life. *Scientific American*, 2007.

[4] A. R. Doherty and A. F. Smeaton. Automatically segmenting lifelog data into events. In *WIAMIS 2008 - Workshop on Image Analysis for Multimedia Interactive Services*, 2008.

[5] A. R. Doherty, A. F. Smeaton, K. Lee, and D. P. Ellis. Multi-modal segmentation of lifelog data. In *RIA0 2007 - Large-Scale Semantic Access to Content (Text, Image, Video and Sound)*, 2007.

[6] G. Gaughan and A. F. Smeaton. Finding new news: Novelty detection in broadcast news. In *AIRS 2005 - Second Asia Information Retrieval Symposium.*, pages 583–588, 2005.

[7] 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 Ubiquitous Computing Conference*, 2006.

[8] W.-H. Lin and A. Hauptmann. Structuring continuous video recordings of everyday life using time-constrained clustering. In *SPIE - Multimedia Content Analysis, Management, and Retrieval*, 2006.

[9] D. Tancharoen and K. Aizawa. Novel concept for video retrieval in life log application. In *PCM Pacific Rim Conference on Multimedia*, 2004.