

# Creating Digital Life Stories through Activity Recognition with Image Filtering

Basel Kikhia, Johan E. Bengtsson, Kåre Synnes,  
Zaheer ul Hussain Sani, and Josef Hallberg

Luleå University of Technology, Sweden  
Department of Computer Science & Electrical Engineering  
{Basel.Kikhia,Johan.E.Bengtsson,Kare.Synnes,Josef.Hallberg}@ltu.se,  
zaheer@memorylane.nu

**Abstract.** This paper presents two algorithms that enables the MemoryLane system to support persons with mild dementia through creation of digital life stories. The MemoryLane system consists of a Logging Kit that captures context and image data, and a Review Client that recognizes activities and enables review of the captured data. The image filtering algorithm is based on image characteristics such as brightness, blurriness and similarity, and is a central component of the Logging Kit. The activity recognition algorithm is based on the captured contextual data together with concepts of persons and places. The initial results indicate that the MemoryLane system is technically feasible and that activity-based creation of digital life stories for persons with mild dementia is possible.

**Keywords:** life logging, life story, memory assistance, activity recognition, image filtering

## 1 Introduction

The idea of creating digital life stories goes back to 1945 when Vannevar Bush envisioned “a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility” [1]. In order to fulfill this vision, life should be logged and clustered in manageable shots/activities [7], so it will be available for reminiscence and later retrieval.

Reviewing images and written diaries are the most obvious examples for stimulating memory with past activities [15]. To support this stimulation through a digital tool, images should be collected during the day then clustered into activities. Therefore, activity recognition is the key component for achieving digital life stories. Moreover, clustering life into activities supports the idea of having a written diary in a digital format. It is also important to filter out images that do not give meaningful information rather than overloading the user with too many images to review.

This paper discusses challenges in implementing activity recognition with image filtering, and it presents techniques that have been implemented in a tool and field tested by 4 researchers and 4 users suffering from mild dementia. The tool logs

personal context and content of a person, then recognizes a sequence of activities based on the visited places and persons met - a digital life story, and finally allows the user to review and adjust the activities and related images, as a reminiscence process and to ensure suitable quality for later context-dependent retrieval. The research questions discussed in this paper are:

1. How can the number of images presented to the user in a life logging system be minimized using filtering techniques?
2. How can activities be automatically recognized based on context data, prior knowledge of places and persons, and a minimal set of equipment?

The rest of this paper is organized as follows: In Section 2, state-of-the-art related work is presented. Section 3 discusses the Logging Kit, presenting image filtering techniques to be embedded within the system, that drastically reduces the number of images for the user to review. Section 4 discusses the Review Client, and it presents the activity recognition techniques and the learning mechanism for increasing the system's understanding of known persons and places. The research questions are directly addressed in section 5, while section 6 concludes the paper and presents future work.

## **2 Related work**

Several techniques have been proposed for an efficient method in recognizing activities. For instance, using context to search parts of the recorded video [12], use of a wearable camera for capturing video and a sensor of brain waves [11], face detection and low level analysis of images [9].

Annotating images with contextual data has been explored by several studies. For instance, authors in [13] annotate an image with the location where it was taken, using a GPS device, and relating digital camera images to locations based on time. Most of the previous work have been using multiple devices and sensors, which results in having many sources of data for building knowledge into the system. Intensive studies have recently been done by using a wearable digital camera called SenseCam [4] [5] [8] [15], that periodically captures fish-eye lens images without user intervention. Onboard sensors in SenseCam are used to trigger additional image capture [15]. SenseCam reduces the need for equipment for a life logging system because of the built-in sensors. Authors in [4] used low-level features of SenseCam images to define high level semantic concepts such as eating, road, sky, office, etc. 27 semantic concepts have been evaluated for their relevance in accurate activity detection. This work provides significant knowledge about how content-based retrieval could be used to support clustering a day into activities. MPEG-7 and SURF have been used in combination with SenseCam device in [5]. The sensor readings were fused and normalized to improve the representation of events. Evaluation showed the possibility of retrieving similar events over a longer time span. For instance, the user can retrieve all the events when she was sitting in a restaurant. The studies in [4] and [5] concentrated on using content-based retrieval to support event segmentation. In [8] the feasibility of using GPS data and Bluetooth MAC addresses to improve retrieval of similar events was successfully tested.

In MemoryLane, we have emulated some of SenseCam functionality by developing software for the Microsoft Windows based mobile device “HTC Touch Cruise”. This device has been used as a logging device worn around the neck to automatically capture images, Bluetooth MAC addresses and GPS data. The captured data are then transferred to a personal computer, where images are filtered automatically. We also extend previous work by introducing a Review Client, that clusters the collected data into activities with identified places and persons that can be adjusted by the user, to semi-automatically form digital life stories.

### **3 Logging Kit**

The Logging Kit consists of two parts: MobileLogger and LogSync. MobileLogger runs in the logging device, and it periodically logs GPS, Bluetooth MAC addresses, and Image data. LogSync is a computer application that transfers the logs from the logging device when it is connected to the computer. It also filters the images while transferring.

The logging device was also given some add-on features; a temporary camera deactivator button and a manual shutter. The camera deactivation button disables image capturing for seven minutes, while playing a configurable voice message “camera deactivated”. The manual shutter button captures an extra image while playing a shutter sound. GPS data and Bluetooth MAC addresses are logged and appended into separate XML files that are used later on for analyzing the day and clustering it into activities.

The set of images is reduced by applying filters for dark and blurry images, and for eliminating sequences of similar images, then the remaining images after filtering are rotated as necessary to compensate for camera orientation.

#### **Dark images**

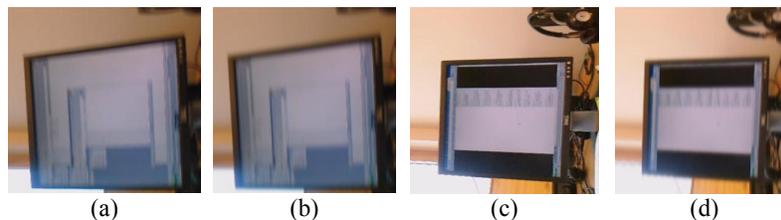
Dark images frequently occur when something obscures the camera lens, or due to low environmental lighting. To filter out dark images, LogSync computes the HSB (Hue, Saturation, Brightness) values of all pixels in an image. The image is considered too dark when the average brightness value is less than a threshold value of 0.2. The threshold value has been chosen based on observation, and it showed positive results in eliminating dark images. Preliminary user feedback confirms the effectiveness of the brightness filter.

#### **Blurry images**

Images captured can be blurry due to many factors, for example movement of objects in view, or movement of the camera itself while capturing. Lower camera resolution and movement of the person who is wearing the logging device results in more blurry images. Blurriness caused by movement is known as motion blur [14] or linear blur [3], and there are many techniques to identify and improve images suffering from these defects [14]. Our approach is to identify blurry images and filter them out without attempting to improve them. This ensures that only non-blurry images are presented to the user, and it also saves processing time.

Motion blur that is caused by horizontal movement of objects can be applied to a non-blurry image by implementing Gaussian blur [2] horizontally. It has been observed that when applying a small amount of Gaussian blur horizontally on an already horizontally blurry image, the image will change slightly. While applying the same amount of Gaussian blur horizontally on a non-blurry image will change the image significantly. Therefore, the adopted approach is to apply little horizontal blur on an image and compare the result with the original. The comparison method described later to check similarity is used here. If the computed similarity is equal or more than 99.7 percent, the image is considered as blurry. The users during the field test complained about having few images by the end of the day, so the value 99.7 should be reduced to make the system results in much more images.

In Fig. 1 below, image (a) is originally captured by the logging device. A small amount of horizontal blur is applied on it, changing it only slightly (image (b)). When applying a small amount of horizontal blur on image (c), the result was image (d). Checking the similarity of (c) and (d) showed a result less than 99.7, so (c) is defined as a non-blurry image, while (a) would be automatically discarded as blurry image.



**Fig. 1.** (a) Blurry image (b) Same image “a” after applying a small amount of horizontal blur. (c) Non-blurry image (d) Same image “c” after applying a small amount of horizontal blur

### Similar images

An activity can have long sequences of similar images. Therefore, efficient methods to check for image similarity are called for. To compare two images, LogSync compares their grayscale form using the MAE (Mean Absolute Error) method which is computationally cheap [10]. Images are considered similar if the MAE value is at least 90 percent.

Groups of similar images are identified with the same group ID, and this is appended into an XML file. The first transferred image is given a “0” similarity percentage (no previous image to compare with). The similarity check is then applied on each pair of consecutive images, and the similarity values are appended into the XML file. If the similarity is less than 90 percent, a new group ID will be assigned to the transferred image. The system will assign the middle image of each group as a representative image of the group. The test during implementation showed that 90 percent is a reasonable value to define similarity between 2 images.

In Fig. 2, images are taken during an activity and checked for similarity. Images “b”, “c”, and “d” are identified similar to “a”.



**Fig. 2.** Images captured during an activity and identified as similar

## 4 Review Client

The Review Client typically runs on a touch screen computer at home. It analyzes the transferred logs in order to cluster them into activities (based on context data), relates content data (images) to those activities, and presents the day to the user as a sequence of activities in chronological order. GPS coordinates are used to detect places, while Bluetooth MAC addresses of nearby mobile devices are used to detect presence of persons. The Review Client basically creates digital life stories describing the life of a person based on visited places and persons met.

### 4.1. Detecting Context

The Review Client is limited by the resolution of the images captured by the logging device, for example creating difficulties in detecting the presence of faces. It is mentioned in [13] that using Bluetooth detection with face detection can improve the detection of persons' presence, but this relies on high quality images. Instead Bluetooth MAC addresses is the only source used to detect the presence of persons. The Bluetooth MAC addresses of mobile devices of known persons were added manually to the system.

Known places were defined using polygons based on GPS coordinates. Some places are contained by other places. For example, a city can contain a neighborhood that contains a shopping mall. In case of overlapping, the system takes the sub-place to give more accuracy about the user location. Partially overlapping places were not evaluated. There are many sources that cause GPS position error such as few satellites, atmospheric delay, receiver noise and receiver clock errors [6]. During field testing, GPS errors occurred when the logging device was indoors with few detected satellites. The field test showed that places should be defined wider to combat GPS errors, and GPS data based on few satellites should be ignored.

### 4.2. Analyzing logs

After transferring the logs from the logging device, the Review Client analyzes the GPS data to identify the known places that the user visited during the day. It also analyzes the Bluetooth data to identify the known persons that the user met. After identifying the places and persons, the Review Client searches for time-based correlation resulting in activities "In a Place", "With a person", or "In a place with a

Person”. In our first implementation, each distinct set of present persons is considered to belong to an activity. For instance, if the user spends time with Carlos and Johan, then Eva joined them later, two activities with two different times will be generated “With Carlos and Johan” and “With Carlos, Johan, and Eva”.

The activity recognition algorithm identifies a single place (or no known place) for an activity, meaning that walking with Johan, for example, from Home to University will be shown as three different activities “Home with Johan”, “With Johan (on the way between home and university)” and “University with Johan”. While the previous example could be shown as “Home and University with Johan”, the priority was given to cluster activities based on places visited rather than on persons met. This means that the Review Client takes all the visited places as a main source to define activities, then relates all known persons who were present in those places during the visit, and finally considers situations when known persons were met in unknown places.

After clustering activities, the Review Client adds the captured images to the activities that coincide with the time when the images were captured. The Review Client presents the activities of the day in chronological order. A subject heading of each activity is generated automatically based on detected context data during the activity (places and persons), and the middle image taken during an activity is proposed as the representative image for the activity. Activities will thus be automatically annotated with information about subject, representative image, place, persons, and captured images.

### **4.3. Adjusting activities**

When reviewing the activities of the day, the user can adjust context and content data or even discard unwanted activities. This will empower the user and improve the accuracy of the saved activities. For instance, the user can change context data, or skip some of the images taken during an activity.

It is also possible to add new context data to the system. In case of adding a new place, the user has to choose an image for that place and enter some relevant information, then the Review Client will search for the closest detected GPS location when that image was captured. Based on that GPS location, a polygon will be created around this point, and the place will become a known place. Adding persons is done in a similar way by choosing an image and entering relevant information for that person. When adding a person, the right Bluetooth MAC address for that person is typically not well-defined since many devices not related to that person could have been detected at that time. A newly added place will therefore be detected if the user visits it again, while a newly added person will not be detected if the user meets this person again. However, newly added persons and places will appear as known context data in the system, and the user can manually relate them to any activity.

## **5 Discussion**

*How can the number of images presented to the user in a life logging system be minimized using filtering techniques?*

Brightness, blurriness, and similarity filters can be embedded within the life logging system to reduce the number of the images, see section 3. Those filters are implemented sequentially to avoid extra operations within the system. The brightness filter will be applied first, discarding all images that are too dark. A blurriness filter is then applied, keeping only images that are of good quality. The remaining images are then compared to each other sequentially and clustered in groups based on similarity, then the system picks up the middle image from each group.

Our implemented approach helps in eliminating bad quality images and avoids keeping similar images. The filters work unobtrusively during transfer of images from the logging device to a personal computer.

*How can activities be automatically recognized based on context data, prior knowledge of places and persons, and a minimal set of equipment?*

Relying solely on context data in recognizing activities enables use of a single logging device. A logging device that periodically logs GPS, Bluetooth MAC addresses, and images data enables identification of places, and persons' present. Digital life stories based on visited places and persons met with associated images might be used to support memories of past activities, or at least make possible to later retrieve of the captured data based on context.

Known places and known persons need to be configured into the system in order to automate the activity recognition, and to reduce intervention by the user side in creating the life story. Having places and sub-places helps specifying the exact place, while ensuring that recognized activities can always be associated at least with the current city. While a-priori knowledge of relevant places for the user is helpful, the system can learn new places as the user adds them relating them to images captured during the visit.

Automation and unobtrusiveness keeps the interaction between the user and the system simple and efficient. However, the Review Client also provides easy-to-use interfaces allowing the user to manually navigate, adjust or discard activities.

Field testing has yielded positive user feedback on the perceived usefulness of a limited implementation of such a system.

## **6 Conclusions and Future Work**

The algorithms for image filtering and activity recognition has been implemented in a prototype system, that successfully captures context and content data and organizes it into activities forming digital life stories. Initial results indicate that the MemoryLane system is technically feasible and that activity-based creation of digital life stories for persons with mild dementia therefore is possible.

The next stage of our work is to improve the detection of persons' presence, relying not only on Bluetooth MAC addresses. Improvements to the clustering algorithm are also needed to increase the accuracy of the system, and to enable the system to automatically recognize new contextual data (places and persons).

## 7 Acknowledgments

This work was supported by the eHealth Innovation Centre (EIC) at Luleå University of Technology and the MemoryLane research project, co-funded by the European Regional Development Fund (Mål-2).

## References

1. Vannevar Bush: As We May Think, *The Atlantic Monthly*, 176(1), 101-108, (1945)
2. FM Waltz, JWV Miller: An efficient algorithm for Gaussian blur using finite-state machines. *SPIE Conf. on Machine Vision Systems for Inspection and Metrology VII*, (1998)
3. H. Tong, M. Li, H. Zhang, C. Zhang: Blur detection for digital images using wavelet transform, in *Proc. of IEEE. Int. Conf. on ICME*, (2004)
4. Byrne D, Doherty A.R., Snoek C.G.M., Jones G.F., and Smeaton A.F.: Validating the Detection of Everyday Concepts in Visual Lifelogs. *SAMT 2008 – 3rd International Conference on Semantic and Digital Media Technologies*, Koblenz, Germany, 3-5 D (2008)
5. Doherty A.R, Ó Conaire C, Blighe M, Smeaton A.F, and O'Connor N.: Combining Image Descriptors to Effectively Retrieve Events from Visual Lifelogs. *MIR 2008 – ACM International Conference on Multimedia Information Retrieval*, Vancouver, Canada, 30-31 (2008)
6. Langley, Richard B.: The GPS Error Budget. *Innovation Column in GPS World magazine*, March issue, pp. 51-56. (1997)
7. W.-H. 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)
8. Daragh Byrne, Barry Lavelle, Aiden R. Doherty, Gareth J.F. Jones, & Alan F. Smeaton: Using Bluetooth & GPS Metadata to Measure Event Similarity in SenseCam Images. *Centre for Digital Video Processing (CDVP) & Adaptive Information Cluster (AIC)*, Dublin City University, Dublin 9, Ireland, (2007)
9. Doherty A.R. and Smeaton A.F.: Combining Face Detection and Novelty to Identify Important Events in a Visual LifeLog. In: *IEEE International Conference on Computer and Information Technology, Workshop on Image- and Video-based Pattern Analysis and Applications*, Sydney, Australia, (2008)
10. YC Lin, SC Tai: Fast Full-Search Block-Matching Algorithm for Motion-Compensated Video Compression. *IEEE Transactions on communications*, (1997)
11. Kiyoharu Aizawa, Ken-Ichiro Ishijima and Makoto Shiina: Summarizing Wearable Video. *Proceedings of ICIP 2001, IEEE*, pp 398-401, (2001)
12. Aizawa, K., Hori, T.: Context-based Video Retrieval System for the Life-log Applications. In: *Proc. of MIR 2003*, pp. 31–38. *ACM, New York*, (2003)
13. Bisht M., Swords D., Quigley A., Gaudin B. and Bennett M.: Context-Coded Memories: "Who, What, Where, When, Why?: Supporting Human Memory with Interactive Systems. *Workshop at the British HCI International Conference*, (2007)
14. Ji, H., Liu, C.: Motion blur identification from image gradients. *IEEE Conference on Computer Vision and Pattern Recognition. CVPR* (2008)
15. Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, & Ken Wood: *SenseCam: A Retrospective Memory Aid*. Microsoft Research, 7 JJ Thomson Avenue, Cambridge, CB3 0FB, UK. (2006)