Considering Documents in Lifelog Information Retrieval

Rashmi Gupta
(Supervised by Cathal Gurrin, Dublin City University)
Insight Centre for Data Analytics
Dublin, Ireland
rashmi.gupta3@mail.dcu.ie

ABSTRACT

Lifelogging is a research topic that is receiving increasing attention and although lifelog research has progressed in recent years, the concept of what represents a document in lifelog retrieval has not yet been sufficiently explored. Hence, the generation of multimodal lifelog documents is a fundamental concept that must be addressed. In this paper, I introduce my general perspective on generating documents in lifelogging and reflect on learnings from collecting multimodal lifelog data from a number of participants in a study on lifelog data organization. In addition, the main motivation behind document generation is proposed and the challenges faced while collecting data and generating documents are discussed in detail. Finally, a process for organizing the documents in lifelog data retrieval is proposed, which I intend to follow in my PhD research.

CCS CONCEPTS

• **Information systems** → **Document structure**; **Multimedia and multimodal retrieval**; *Information extraction*;

KEYWORDS

Lifelogging; Lifelog Documents; Information Retrieval; Moment; Activity; Event

ACM Reference Format:

Rashmi Gupta. 2018. Considering Documents in Lifelog Information Retrieval. In *ICMR'18:2018 International Conference on Multimedia Retrieval, June 11–14, 2018, Yokohama, Japan.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3206025.3206081

1 INTRODUCTION

Lifelogging refers to the process of passively acquiring a rich multimodal record (log) of daily life activities using one, or more wearable devices, typically involving a wearable camera [4]. This data is usually referred to as a lifelog and it offers considerable potential to enhance the life experience of the wearer, offering new sources of information for self-knowledge and insight generation [11]. In order to provide useful and indexable content for a lifelog retrieval system, it is necessary to align and synchronize the streams of multimodal lifelog data and combine them into some form of retrievable units. Heretofore, there has been little consideration of what a document means for lifelog-based retrieval. Hence, in this work, I propose a new unit of retrieval for lifelog data, which is an indexable unit of an individual's life experience called an activity. Since lifelogging is an inherently multimodal data capture activity, such an indexable unit would manifest as a fusion of a number of synchronized lifelog data sources at a given point in time, such as audio, video or images from wearable cameras, readings from biometric sensors, communication records, information creation or access logs, and various other sensor sources. Such a unit would represent a contiguous period of time and would form the basis of a document of life experience data. Providing access to a large number of such documents would produce a lifelog retrieval engine.

Although the concept of a document has not been explored in depth in the research field, the idea of lifelog annotation and search has been receiving increasing research attention, with dedicated workshops and collaborative bench-marking efforts underway [9]. Thus far in lifelog research, the unit of retrieval has been the life event [11], the minute as a time-unit [9], or the individual data points (e.g. image, temperature, location)[7]. The contribution of this paper, and of my proposed PhD research is to consider what exactly is a document for lifelog retrieval and how can such documents be generated from continuous lifelog data streams. It is my conjecture that this is not a simple research question, since there are many different approaches that one can take, which can either generate indexable documents at indexing time or form query-specific units in some post-query process. It is these factors that motivate my PhD research at Dublin City University under the supervision of Cathal Gurrin.

2 HISTORY AND BACKGROUND

Gathering data in lifelogging has long history, Richard Buckminster Fuller's Dymaxion Chronofile, which was started in the late 1920s [13] included a complete record of his personal and business data, thousand of physical papers, thousand of hours of audios and videos, hundreds of models and artifacts, 1400 feet of content and seventeen hundred hours of recordings. In this case the unit of retrieval would not be in question since the Dymaxion Chronofile was essentially a collection of physical documents and and objects. Decades later, in 2009 [16], Gordon Bell and Jim Gammell coined the concept "total capture" of personally relevant information in their seminal lifelogging work called MyLifeBits [7]. All this data was indexed into a database and made available through an interface that allowed for the retrieval of any unit of information based on a user query. This was instantiated as the MyLifeBits [7] framework which was the first true lifelog retrieval engine. However in this work, the retrieval challenge is viewed as a data retrieval challenge, rather than an information retrieval challenge and the unit of retrieval was the individual unit of data stored. It is my conjecture that a

ICMR'18, June 11–14, 2018, Yokohama, Japan

ACM ISBN 978-1-4503-5046-4/18/06…$15.00
https://doi.org/10.1145/3206025.3206081

lifelog retrieval system should support a user by returning a ranked list of life experiences, as opposed to data values, which would be necessary for lifelog retrieval due to the vast volumes of data that will likely be contained in lifelogs.

In terms of fusing lifelog data from multiple sources into meaningful units, heretofore, there have been two approaches. The simple approach taken by the NTCIR12-Lifelog collaborative benchmarking exercise was to define a retrieval unit as one minute long [9]. While this is effective when developing and evaluating multimodal ranking models, it is not going to be an effective process for a user wishing to retrieve a past life experience due to the volume of documents that are likely to be considered potentially relevant for any given user need. Another approach is the idea of event segmentation, which segments the continual lifelog data into a set of discrete events, by identifying changes in a user's environment as an indicator of life event changes. Doherty et al. in 2007, gathered five sources of lifelog data such as audio data via audio recorders, continual capture images, temperature, white-light level and accelerometer sensors to provide a first generation human memory augmentation tool based on indexing events. They found that the fusion of image, light and accelerometer sensor worked well for identifying the event boundaries when lifelogger moves from one location to another while image and light helps to identify the activity performed by the lifelogger within same location. Thus, they automatically segmented this collection of personal data into specific events [6]. In addition, Li et al. in 2013, collected a fusion of multi-sensor data as a document and segmented whole day activities into specific chunks of time periods for lifestyle evaluation and retrieval purposes [17]. These event segmentation approaches typically produced a small number of daily events (typically 30-40). This averages to about 20 minutes per event, which for some use-cases is appropriate, but for many, it is likely to be too large or certainly is unlikely to be flexible enough for use in a lifelog retrieval system.

In parallel to these event segmentation approaches, the first collaborative benchmarking exercises were being prepared and have been released in recent years under the umbrella of NTCIR [10] and ImageCLEF. These activities release test collections that provide researchers with access to archives of anonymised lifelog data with associated real-world information needs of lifeloggers [9]. Such datasets give us an indication of the types of queries that a lifelogger will make against their collection, and help to motivate and guide this research.

Besides test collections, other researchers have considered the reasons for accessing lifelogs. I propose that it is necessary to take a step back and explore the many potential ways that lifelogs can be used by an individual in daily life. Sellen and Whittaker [16] suggested the five R's of memory access as reasons why individuals access their memories. By extrapolating from these 5Rs, I can then identify five reasons why an individual might keep a lifelog, which are *Recalling* and *Recollecting* of the past experience to support memory-augmentation applications, *Reminiscing* about past experiences to support personal wellness, *Reflecting* on past life experiences for self-enhancement and finally *Remembering Intentions* which is a means of context-aware memory-assistance.

3 RESEARCH FOCUS

Specifically in this research, I will be focusing on recalling and recollecting as the most appropriate forms of retrieval from lifelogs. Consequently, it is my conjecture that neither minute-based retrieval or event segmentation are suitable to provide a flexible retrieval system for the user which can address many of the five R's of memory access. Defining the minute as the unit is clearly too small to represent a human activity, whereas defining the unit as an event could result in units of retrieval that are too long to adequately address an information need. For example, consider an event that is a party; there could be many user activities contained in one large event. Consequently, in my research, I am proposing a new concept called an activity, which I propose to be a more suitable unit of retrieval. The activity unit is shorter than an event and more meaningful in terms of user activity than a minute.

4 CONSIDERING ACTIVITIES AS A UNIT OF RETRIEVAL

As introduced above, the main motivation for this work is to identify what a lifelog document should be and how to generate such document units from the continuous streams of lifelog data coming off various sensors and wearable devices. Some research has focused on an activity-based measure as the important or indexable unit. Braber used a person's daily food consumption record and environment quality, personal behaviour and biometric data and individual performance in terms of mental and physical activities as a document segmentation model in order to promote health and wellness [3]. Zhen et al. in 2013, collected a fusion of multi-sensor data as a document and segmented daily activities into specific segments for lifestyle evaluation and retrieval purposes [17]. Such a viewpoint puts the individual at the centre of the segmentation process and it is my belief that an activity-based segmentation at indexing time will be the most appropriate query-agnostic unit for lifelog search and retrieval .

In order to understand the context in which I define different units of retrieval, I present a hierarchical model of lifelog data units: **Item:** Item is defined as the smallest retrievable unit, the atomic unit of data, such as an image, temperature reading, location, etc. It is the unit of retrieval that was favoured in MyLifeBits [7]. **Moment:** Moment is defined as a fixed length temporal unit, which heretofore has been a minute. Consequently, there are 1440 moments in each day and the minute unit is represented by a combination of all the atomic items that take place within that minute. Moments were used as the retrieval unit in the NTCIR Lifelog comparative benchmarking exercises [10]. **Activity:** Activity is defined as an un-interrupted sequential state of the individual in terms of their person or environment or stimuli. The activity is the indexing-time unit of retrieval that we define in this work and propose as the most appropriate indexing time unit. It represents a combination of sequential items whose size is dependent on the activities of the individual. **Event:** Event is combination of moments or activities or experiences developed (up until now) at indexing time and typically an event is a longest unit of which there may be 2-4 in any given hour (based on past research). The event has been considered to be the first unit of retrieval for lifelog data and was employed manually in the initial Sensecam image viewer tool [7] as well as the early

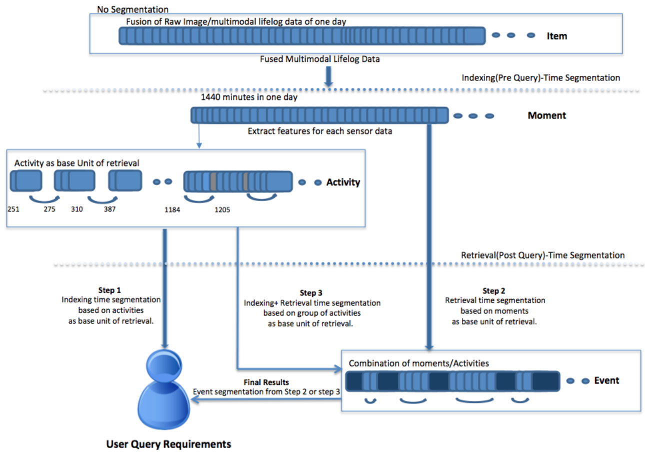

Figure 1: The concept of segmenting one day lifelog visual/sensor archive into events.

work of Doherty et al. in the development of early lifelog search engines [6]. A visualization of numerous different approaches to segmentation of lifelog data is shown in Figure 1. These can be seen as either begin pre-indexing time or post-query time processes. During this research, I intend to explore both pre-indexing and post-query document segmentation (or generation) approaches. In this paper, I report on initial experiments that evaluate the activity-based pre-indexing segmentation (Step 1 in Figure 1). Of course, there are alternative segmentation approaches, such as post-query generation of query-focused temporal segments of lifelog data that are generated in response to a query (Step 2 in Figure 1), or the process of generating events from sequential query-relevant activities in a fused pre/post approach (Step 3 in Figure 1). These alternative approaches will be evaluated over existing datasets in later work.

In terms of activity-based indexing, human activities labeling can range from the broad physical activity measurements (e.g. walking, sitting, running), to detailed activities such as looking, speaking, opening a door, etc. For this work, what I have done is to define a set of sixteen lifestyle activities that are based on, but not the same as Kahenmann's set of the sixteen most enjoyable lifestyle activities [15], which were selected based on personal interview and have been previously used as target activity labels in the SenseSeer smartphone-based lifelogging framework [2]. In this research, activities are defined after examining the long-term lifelogs of two individuals and as such, represent an achievable segmentation target. These sixteen lifestyle activities are detailed in [10] and are designed to explore knowledge mining and visualization of lifelogs.

5 AN APPROACH FOR GENERATING ACTIVITY-BASED LIFELOG DOCUMENTS

Based on my experience of developing and evaluating event-segmentation approaches to lifelog data retrieval, I have created a three-step process for generating activity-based documents for indexing lifelog content. The initial step is to gather sufficient multimodal data from numerous sources, as discussed in [11]. I then propose that these multimodal sources of data are time-aligned, processed to ensure adherence to any legal or ethical data governance expectations (e.g. face blurring or other anonymisation procedures) and then indexable features extracted from the raw lifelog data (e.g. visual concepts from the image content). Following that I propose the fusion of these different data sources into vectors that represent every minute and to simplify the process, the subsequent combination of three minutes into our activity-based segmentation. My proposed process of generating lifelog document is discussed in detail and shown in Figure 2 below.

Data Gathering involves the utilization of multiple sources of unstructured, continuous, raw lifelog data from wearable, fixed and informational (software/online) sensors. Although the data sources are many and varied, I utilize the following sources: **Wearable multimedia information:** Visual and audio data gathered using wearable cameras and microphones, which generates large volumes of content per day (1,500-2,000 images and hours of audio). These give details of the actual human activities and act as reminders for the individual of what they were doing at any point in time. **Human biometrics:** Using a smartwatch we can obtain all-day human biometrics, such as heart rate, calorie burn, and steps, etc. **Human activity:** Logging an individual's physical activities (e.g. walking, driving, running, resting) can be done using wearable devices or smartphone apps , without requiring any user intervention). **Information access:** Using the LoggerMan[12] application, I can collect human-computer interaction data produced during normal computer usage. LoggerMan allows the gathering of a wide range of keyboard, mouse and screen actions, thereby capturing the information creation and consumption activities of the individual.

Following data gathering, I consider the preprocessing of the data that temporally aligns and synchronizes the continual data streams. The multimodal lifelog data gathered by lifeloggers varies in volume, fidelity, accuracy, semantic meaning, and ultimately these vary from one lifelogger to another. Following alignment, it may be necessary to engage in a process of data anonymisation, and many anonymisation techniques can be applied, such as face or tattoo blurring, named identity removal, document/screen blurring from wearable camera data, etc. Finally the preprocessing stage requires feature extraction to convert the raw lifelog data into machine-indexable content. There are many options for selecting

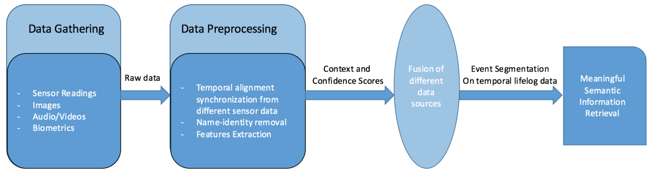

Figure 2: Process of segmenting lifelog data into document units.

sources of visual features for image data, with the most well known being the open-source CAFFE framework[14] that provides a visual annotation of each image from a 1,000 visual feature set.

This data is then fused together to form moments of length one minute before being combined into the indexable activity-based documents by employing a activity recognition algorithm. In my initial work, The two pre-existing segmentation approaches based on MPEG-7 Descriptors implemented by Doherty [5] was re-implemented as a baseline approach and two new approaches (Caffe concepts [14] and Microsoft Computer Vision API image categorization [1]) to segment one day lifelog data into activities were implemented that intend to perform better from baseline approach. Then, euclidean distance measure was implemented to identify the distance between vectors representing the content of every minute. The specific document boundaries were identified using thresholding which was defined based on a manually generated groundtruth by ten users and the average performance of a document segmentation approach using this process was evaluated in terms of precision and recall as shown in Table 1. In results, document segmentation based on image categorization via Computer Vision API approach proved the best approach for document segmentation; for full details see [8].

Event Segmentation Approaches	Threshold Value	Precision	Recall	F1- Score	MCC
Hearst's TexTiling based on MPEG-7 Descriptors	mean(K = 0.5)	20.6	65.4	30.7	6.98
	Kapur	**20.6**	**65.8**	**31.3**	**7.42**
Non TexTiling based on MPEG-7 Descriptors	**mean(K = 0.5)**	**29.5**	**60.6**	**38.7**	**22.0**
	Kapur	29.4	60	38.7	21.43
Caffe Visual Concepts	**0.4**	**70.4**	**72**	**69.3**	**64.3**
	0.5	64.6	76.5	68.5	62.9
	0.6	56.4	80.9	64.8	58.6
	0.7	40.5	88.2	54.2	46
Image categorization via MS Concepts	0.4	78.3	65.5	69.2	65.2
	0.5	77.5	66.2	69.4	65.4
	0.6	77.2	67.2	69.8	61.4
	0.7	**76.2**	**68.3**	**70.1**	**65.7**
Summary Results (Overall Best Approach)	**Threshold Value**	**Precision**	**Recall**	**F1- Score**	**MCC**
MS Concepts	**0.7**	**76.2**	**68.3**	**70.1**	**65.7**
Caffe Concepts	0.4	70.4	72	69.3	64.3
MPEG-7 without TexTiling	mean (K = 0.5)	29.5	60.6	38.7	22.0
MPEG-7 with TexTiling	Kapur	20.6	65.8	31.3	7.42

Table 1: Various Document Segmentation Approaches and found Overall Best Approach.

6 CONCLUSIONS AND FUTURE WORK

In this paper, I motivated the need for generating documents in lifelogging based on activities as a basic unit of retrieval. I identified a taxonomy of lifelog documents from the single data unit to the event. Finally, I proposed a process through-which lifelog documents can be generated.

In future work, I can identify the following tasks that I need to complete: Engage in a larger-scale study on the effectiveness of activity-based segmentation for lifelog retrieval; Develop an approach to query-time document-generation from sequential moments that are highly-ranked in relation to a given information need. This approach will be compared to the activity-based segmentation; Develop a third approach (query-biased combination of highly-ranked activities) to provide a third alternative segmentation approach; Compare the three approaches to segmentation for the processing of known-item lifelog queries.

ACKNOWLEDGMENTS

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289.

REFERENCES

[1] [n. d.]. Microsoft's Computer Vision API. https://docs.microsoft.com/en-us/azure/cognitive-services/computer-vision/category-taxonomy.
[2] Rami Albatal, Cathal Gurrin, Jiang Zhou, Yang Yang, Denise Carthy, and Na Li. 2013. Senseseer mobile-cloud-based Lifelogging framework. In *ISTAS*. IEEE, 144–146.
[3] Maarten den Braber. 2016. The Emergence of Quantified Self as a Data-driven Movement to Promote Health and Wellness. *Proceedings of the first Workshop on Lifelogging Tools and Applications* (Oct. 2016).
[4] Martin Dodge and Rob Kitchin. 2007. 'Outlines of a world coming into existence': Pervasive computing and the ethics of forgetting. *Environment and Planning B: Planning and Design* 34, 3 (2007), 431–445. https://doi.org/10.1068/b32041t
[5] Aiden R. Doherty and Alan F. Smeaton. [n. d.]. Automatically segmenting lifelog data into events. *9th International Workshop on Image Analysis for Multimedia Interactive Services* (june [n. d.]).
[6] Aiden R. Doherty, Alan F. Smeaton, Keansub Lee, and Danial P.W. Ellis. 2007. Multimodal segmentation of lifelog data. *Large Scale Semantic Access to Content (Text, Image, Video, and Sound)* (june 2007), 21–38.
[7] J. Gemmell, G. Bell, and R. Lueder. 2006. MyLifeBits: a personal database for everything. *Commun. ACM* 49, 1 (2006), 88–95.
[8] Rashmi Gupta and Cathal Gurrin. 2018. Approaches for Event Segmentation of Visual Lifelog Data. *24th International Conference on Advances in Multimedia Modeling, MMM 2018* (2018).
[9] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatal. 2016. NTCIR Lifelog: the first test collection for lifelog research. In *SIGIR'16 - the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 705–708.
[10] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Rashmi Gupta, Rami Albatal, and Duc Tien Dang Nguyen. 2017. Overview of NTCIR-13 Lifelog-2 Task. *The Thirteenth NTCIR conference (NTCIR-13)* (dec 2017).
[11] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. 2014. LifeLogging: Personal Big Data. *Foundations and Trends in Information Retrieval* 8, 1 (2014), 1–125. https://doi.org/10.1561/1500000033
[12] Zaher Hinbarji, Rami Albatal, Noel E. OâĂŹConnor, and Cathal Gurrin. 2016. LoggerMan, a comprehensive logging and visualisation tool to capture computer usage. *22st International Conference on MultiMedia Modelling* (2016), 342–347.
[13] Roberto G. Trujillo Hsiao-yun Chu. 2009. New Views on R. Buckminster Fuller. *New Views on R. Buckminster Fuller* (2009), 6–23.
[14] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. *arXiv preprint arXiv:1408.5093* (2014).
[15] Daniel Kahneman, Alan B. Krueger, David A. Schkade, Norbert Schwarz, and Arthur A. Stone. 2004. A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science* 306 (2004), 1776–1780.
[16] Abigail Sellen and Steve Whittaker. 2010. Beyond total capture: A constructive critique of lifelogging. *Commun. ACM* 53 (may 2010), 70–77.
[17] Li Zhen, Wei Zhiqiang, Jia Wenyan, and Sun Mingui. 2013. Daily Life Event Segmentation for Lifestyle Evaluation Based on Multi-Sensor Data Recorded by a Wearable Device. *Conf Proc IEEE Eng Med Biol Soc* (oct 2013).