

Event Based Information Retrieval from Digital Lifelogs

Rashmi Gupta

B.Sc, M.Sc

A Dissertation submitted in fulfilment of the
requirements for the award of
Doctor of Philosophy (Ph.D.)

to the



Dublin City University

School of Computing

Supervisor: Dr. Cathal Gurrin

September 2020

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Ph.D is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Sign:

Student No.: 15213039

(Rashmi Gupta)

Date:

Acknowledgements

I express my heartfelt gratitude to my supervisor, Dr. Cathal Gurrin, for giving me an excellent opportunity to work under his supervision, for supporting my research directions and encouraging me throughout my PhD, and for being my true inspiration in every aspect of life. I will be ever grateful to my supervisor Dr. Mark Roantree for his consistent encouragement and motivation. I would like to acknowledge Professor Rune Sa  tre, NTNU Trondheim, Norway for motivating and supporting me to pursue my PhD.

I extend my sincere thanks to the Insight SFI Research Centre for Data Analytics and Science Foundation Ireland for funding my research.

I thank all my friends and colleagues in DCU, for their companionship, for sharing the lighter moments and including me in all their celebrations. I would like to express the deep love for my friend Koel Dutta Chowdhury for her continuous support and motivation.

On a personal note, I thank my parents (Raj Kumar Gupta and Seema Gupta), my elder sister (Dr. Mayuri Gupta), brother-in-law (Sushil Singh), my younger brother (Suraj Gupta), and love of my life (Achint Atri) for being my pillars of strength and support.

I would like to extend my heartfelt thanks to God who provided me with inner strength, self-esteem, and loads of patience in each phase of my PhD.



Contents

1	Introduction	1
1.1	Chapter Overview	1
1.2	Introduction to Lifelogging	1
1.2.1	Lifelog as an IR Challenge	3
1.3	Research Motivation	5
1.4	Research Challenges	8
1.5	Hypothesis and Research Questions	10
1.6	Contributions of this Dissertation	14
1.7	Production List	14
1.7.1	List of Poster Presentations	16
1.8	Dissertation Outline	16
2	Related Work	20
2.1	Chapter Overview	20
2.2	Evolution of Lifelogging	20
2.3	Data Capture in Lifelogging	22
2.3.1	Raw Data Sources	23
2.3.2	Descriptive Data Sources	27
2.4	Information Retrieval from Multimodal Lifelogs	31
2.4.1	IR Approach	33
2.4.2	Evaluation of Lifelog Data Access	35
2.5	Event Segmentation of Lifelog Data	38
2.5.1	Approaches to Static Event Segmentation	39
2.5.1.1	Conclusion to this Review	44

2.6	Chapter Summary	45
3	Research Methodology and Research Configuration	47
3.1	Chapter Overview	47
3.2	Lifelog Data Collections	48
3.2.1	Challenges of Building Test Collection	50
3.2.2	Lifelog Data Sources Used in this Research	51
3.3	Users and User Issues	53
3.3.1	User Selection	53
3.3.2	User Preparation	54
3.3.3	User Interrogation	54
3.3.4	Data Governance	54
3.3.4.1	Data Protection	55
3.3.4.2	Data and Ethics	55
3.3.5	Experimental Configuration	57
3.3.6	Results and Findings	60
3.3.6.1	Stage 1: Reasons for participation in user study . .	60
3.3.6.2	Stage 2: Three different privacy levels	61
3.3.6.3	Stage 3: Experiences of the wearers while lifelogging	66
3.3.7	List of Guidelines to Preserve the Anonymity	67
3.4	Releasing Privacy-Aware Lifelog Test Collections	68
3.4.1	Principles of Generating Lifelog Dataset	69
3.4.2	Process of Releasing Lifelog Test Collections	72
3.4.3	Lifelog Test Collections Used in this Research	74
3.5	Basic Units of Retrieval	75
3.6	Evaluation Measures and Techniques	78
3.6.1	Evaluation Metrics	78
3.7	Summary of Proposed Approaches	80
3.8	Chapter Summary	81

4	Lifelog Data Processing and Data Analytics Techniques	82
4.1	Introduction	82
4.2	Lifelog Data Preprocessing Techniques	82
4.2.1	Checking the Shape of Lifelog Images	83
4.2.2	Resizing Lifelog Images	83
4.2.3	Removing Noise from Lifelog Images	84
4.2.4	Feature Extraction and Concept Detection	84
4.2.4.1	Automatic Concept Detection Tools	85
4.2.4.2	Manual Annotations	87
4.2.4.3	Contextual and Descriptive Metadata	88
4.3	Lifelog Data Analytics Techniques	89
4.3.1	Vector Distance Methods	89
4.3.2	Thresholding Techniques	90
4.4	Chapter Summary	91
5	Static Segmentation of Lifelog Data	92
5.1	Introduction	92
5.2	Events are Documents	92
5.3	Event Segmentation of Lifelog Data	95
5.3.1	Baseline Segmentation Approach to Event Segmentation	98
5.3.2	Event Segmentation Based on Visual Concepts	99
5.3.3	Event Segmentation Based on Image Categories	100
5.3.4	Distance Measure	101
5.3.5	Threshold Determination	101
5.3.6	Avoiding Over-Segmentation	102
5.4	Evaluation	102
5.4.1	Initial Dataset	103
5.4.2	Evaluation Methodology	103
5.4.3	Results and Discussion	105

5.5	Multimodal Lifelog Segmentation Based on Activity Retrieval Unit .	109
5.5.1	Lifelog Activity Segmentation	110
5.5.2	Dataset and Groundtruth Generation	111
5.5.3	Evaluation Methodology	112
5.5.4	Results	113
5.6	Chapter Summary	116
6	Dynamic Segmentation of Multimodal Lifelog Data	117
6.1	Introduction	117
6.2	Dynamic Activity based Segmentation of Multimodal Lifelog Data .	119
6.2.1	IR Experimental Methodology	120
6.2.2	User Query Topics	122
6.2.3	Evaluation Methodology	123
6.2.3.1	Results and Discussion	123
6.3	Considering Manual Annotations	125
6.3.1	Automatic Annotation Based Segmentation	125
6.3.2	Manual Annotation Based Segmentation	126
6.3.3	Fused Annotation Based Segmentation	127
6.3.4	Developing New User Query Topics	128
6.3.5	Evaluation Methodology (Precision Focused)	128
6.3.6	Results and Discussion	130
6.4	Identifying the Best Segmentation Approach	133
6.4.1	Exploring the Best Segmentation Algorithm	134
6.4.2	Discussion	136
6.5	Chapter Summary	137
7	Conclusion	138
7.1	Dissertation Overview	138
7.2	Research Questions and Proposed Solutions	141
7.3	Contributions of this Dissertation	143

7.4	Limitations of this Work	144
7.5	Suggestions for the Future Research	145
7.6	Concluding Remarks	147
	Bibliography	148
	A Appendix	167
A.1	Lifelog Event Segmentation Task Topics	168
A.2	Ontology to Annotate Lifelog Data	174
A.3	LADT - Lifelog Activity Detection Task Topics	176
A.4	Appendix to Publications	177

List of Tables

2.1	The summary details of query topics associated with specific lifelog tasks for publicly available lifelog test collections at NTCIR and LSC workshop.	38
2.2	Summary of lifelog data sources and associated content retrieved based on different scales or units such as full document, atomic unit (au), temporal unit and event segmentation.	45
3.2	Stage 1: Summary results of reasons for being participant in user study.	61
3.3	Summary of the privacy assessment from the 25 participants of various types of content in visual lifelog data across three access levels: <i>Private</i> , <i>Semi-Private</i> , and <i>Public</i> , sorted in decreasing order of privacy concern.	62
3.4	The summary details of all the publicly available datasets along with other metadata being used for this research.	74
3.5	Summary of experimental configuration for this research.	81
5.1	The summary information of participants and their count of segmented ground-truth events.	103
5.2	Overall Thresholding Performance based on MPEG - 7 Descriptors [1].	106
5.3	Overall Thresholding Performance based on visual concepts using Caffe framework [2].	106
5.4	Overall Thresholding Performance based on image categories using Microsoft cognitive service API [3].	107
5.5	Overall thresholding performance based on MPEG - 7 descriptors [1], Caffe visual concepts [2], image categorisation via MS Concepts [3]. .	107

5.6	Various evaluation approaches implemented to calculate the accuracy of static (indexing-time) activity boundary detection and segmentation using image categorisation from MS visual concepts [3] and identify the best evaluation approach.	114
6.1	Dynamic event segmentation based on activity or group of activities as basic unit of retrieval (Dynamic Approach in Figure 6.1).	124
6.2	List of 12 broad focused and 12 narrow focused user queries (topic titles only).	129
6.3	Performance (i.e. precision) of the dynamic lifelog segmentation based on manual annotations and automatic visual concepts.	131
6.4	Comparing system performance (i.e. average precision) of query-specific dynamic event segmentation based on three different approaches to annotation.	132
6.5	Accuracy scores for each activity in dynamic Segmentation approach.	135
6.6	Performance of the static and the dynamic segmentation approaches.	136

List of Figures

1.1	A Lifelogger with her daily wearable sensors: smartwatch (on wrist), Autographer camera (wearing around her neck), ion/snapcam camera (attached with pocket).	3
1.2	Example of different types of lifelog events and the transition of these variable-sized lifelog events in a day.	7
1.3	The Structure of this Dissertation.	17
3.1	The experimental methodology for information retrieval in lifelog data.	48
3.2	Example images from the specified three categories (i.e A, B and C) of anonymisation & highlight potential differences in willingness to share data i.e difference in ratio of bedroom images messiness from Access Category A: <i>Private</i> with Access Category B: <i>Family/ friends</i>	65
3.3	A lifelogger with his everyday wearable devices: A Narrative camera (wearing on his neck) and a smartwatch (the black watch on his right arm).	73
3.4	Hierarchical Model to define basic units of retrieval in lifelog data based on the intentions of recalling and recollecting personal life experience.	76
3.5	Confusion matrix: A standard metric for segmentation and classification problems.	79
4.1	Pseudo-code for lifelog data pre-processing.	84
4.2	Example of lifelog images with extracted visual features associated with confidence values.	86

4.3	Example of visual features extracted using vision recognition and content detection API [3].	87
4.4	Example image with extracted automatic visual features, manual annotations, and metadata.	89
4.5	Pseudo-code for calculating the Euclidean vector distance method and mean thresholding method.	91
5.1	Process of generating lifelog documents in lifelogging.	93
5.2	Example of segmented daily life user events in lifelog dataset using Caffe framework [2] and Microsoft's cognitive service [3] concept detector tools.	96
5.3	Example of signifying the transition between different user events.	96
5.4	Methodology to segment one day visual lifelog data into events at indexing time.	97
5.5	Lifelog event segmentation based on Caffe visual concepts of one day data [2].	99
5.6	Algorithm for segmenting visual lifelog data using visual concepts from Caffe framework [2].	100
5.7	Example of segmented lifelog dataset based on categories [3].	100
5.8	Algorithm for reading lifelog data and visual concepts from Microsoft cognitive services API.	101
6.1	Working on activity based dynamic segmentation of lifelog data.	118
6.2	Example of top ranked query-focused retrieved activities i.e. four different visual shots of detection and segmentation typically belong to one specific activity but four different day environments e.g. Query: find where the user is using digital screens.	120
6.3	Process of identifying effective top-ranked activity boundaries in LSC Dataset.	121

6.4	Visual representation of identified human activity such as find the time where an individual is using digital screens in LSC dataset. . .	122
6.5	Considering manual annotations in activity based dynamic segmentation of multimodal lifelog data.	126
6.6	Examples of automatic high-quality visual concepts from concept detector using Microsoft cognitive services API [3].	127
6.7	Example of manual annotations based on 24 real-world life activities.	127
6.8	The sliding window approach for the evaluation of activity detection and segmentation algorithm.	133
6.9	Identifying an optimal window size (i.e. +/-[1 to 5] minutes) in sliding window algorithm.	134

Event Based Information Retrieval from Digital Lifelogs

Rashmi Gupta

Abstract

For the increasing prevalence of large volumes of continuous multi-sensor personal data, new challenges have aroused such as multi-sensor data collection, lifelog data storage, meaningful information retrieval, and data representation across different retrieval systems. It leads a proper consideration of the segmentation of such large lifelogs into manageable discrete units as per the user information requirements. This research investigates the approaches to address the above challenges. The goal is to offer a novel and effective lifelog segmentation approaches to ad hoc search through lifelog archives. Instead of the conventional static indexing-time based segmentation of lifelog data, we proposed dynamic query-time based segmentation of multimodal lifelog data that we believe better satisfies the user information requirements.

The main contributions of this dissertation are: First, identifying the process of constructing privacy-aware comparative reusable lifelog test collections including various challenges and principles, We introduce user's issues such as user selection, user interrogation, data governance, data protection and ethics to test segmentation, and remembered information access; Second, defining the concept of documents in multimodal lifelog data retrieval; Third, defining the baseline segmentation approach to segment visual lifelog data, and developing state-of-the-art segmentation approach for multimodal lifelog data; Fourth, introducing a new dynamic query-time based segmentation algorithm for multimodal lifelog data.

A publicly available dataset lifelog search challenge (LSC2018) is used for the comparative evaluation. We used different broad and narrow focused user queries, with promising results that outperform existing conventional static segmentation approaches. In summary, this dissertation primarily contributes to the field by

identifying a new segmentation approach i.e. indexing-time or query-time to extract the meaningful information from multimodal lifelog data and highlights several potential topics of interest for the research community.

Chapter 1

Introduction

1.1 Chapter Overview

This chapter provides an overview of the research work in this dissertation. It outlays motivation and challenges of the research, proposes hypothesis and formulates the research questions. Section 1.2 introduces lifelogging, wearable devices used to collect multi-sensor lifelog data and the general overview of analysing multimodal lifelog data. Section 1.3 discusses the specific motivations underpinning this research. Section 1.4 poses the challenge that needs to be solved in this research work. Section 1.5 presents the hypothesis and research questions derived from the previously reported work on event boundary detection and segmentation of lifelogging. Section 1.6 list the contributions of this dissertation. Section 1.7 list all the publications and section 1.8 outlines the organisation and structure of this dissertation.

1.2 Introduction to Lifelogging

Lifelogging is defined as digital archiving of rich personal life experience [4]. It is a relatively new phenomenon in which individuals (the *lifelogger*) track and record their daily life experience for a variety of purposes such as dietary tracking, sleeping habits, exercise routines and/or social interactions. This is carried out with the help of one or more lifelog devices. There is no universal standard definition of lifelogging. We have chosen the definition of lifelogging as defined by

Dodge and Kitchin (2007) [5] as “a form of pervasive computing, comprising a unified digital record of the totality of an individual’s experiences, captured multi-modally through digital sensors and stored permanently as a personal multimedia archive”. Such individuals, who are also called lifeloggers, gather this multimodal lifelog data. These lifeloggers could be researchers or any individual who wishes to capture a detailed digital trace of their life [4]. Lifelogging provides detailed information about daily activities of an individual and could help to change in behaviour. This can lead to positive impact in an individual’s life.

The lifelog data could be a collection of images, audios, videos, text data or biometric data. This data can be highly sensitive and private to the lifelogger. The data is gathered by using various small wearable devices and software. Cost reductions have led to an increased availability of wearable devices for lifelogging (shown in Figure 1.1). An example here is Microsoft SenseCam [6] used in MyLifeBits project [7]. This wearable camera captures the field of view of the wearer and provides a detailed record (*log*) of the life of an individual. It can capture about 2000 images autonomously per day without human intervention. A smart watch can monitor biometrics of an individual and it can capture additional data about the individual’s activities. It can be used by interested consumers self¹ community [8]. In addition, smartphones carry a range of sensors such as accelerometers, gyroscope and capture devices such as microphones, cameras, which can help in lifelogging.

Following this former work of digitally record daily life, we report that most popular wearable devices used in lifelogging are wearable cameras (e.g. autographer or narrative clip); sensors (e.g. audio/video recorders); fitness trackers (e.g. fitbit or apple watch); physical activity sensors (e.g. moves app and gyroscope); and human computer interaction systems (e.g. loggerman²) [9]. These can create digital diary logging individual’s life activities. A rich personal lifelog data can be captured continuously and passively using these lifelogging devices.

¹<http://quantifiedself.com/>

²<http://loggerman.org/>



Figure 1.1: A Lifelogger with her daily wearable sensors: smartwatch (on wrist), Autographer camera (wearing around her neck), ion/snapcam camera (attached with pocket).

Thus, some aspect of quantified self where capturing the detailed record (log) of the individual's life activities by an individual is becoming a normative activity [4].

1.2.1 Lifelog as an IR Challenge

One of the major challenges of any digital archive is to support and organise efficient access. Building an index over lifelog data is required to support queries. Life log data can range above millions of records and without appropriate an index it can quickly become unmanageable. Therefore, there is a challenge of segmenting this continuous stream of lifelog data into most appropriate index for analysis, retrieval, and presentation. We applied organisational methodology based on the core concept of a type of document such as web search, photo retrieval, and/or video search as an indexing unit. A document is the atomic unit of retrieval in conventional search and retrieval method. A document could be considered as a basic unit in case of a web search, but for other information retrieval problems, a document can be too large to be of any value for the user. Consider video search for example, a multi-hour video can be seen as too large to be useful, and thus, the video is often segmented into sub-units called video shots or scenes, which are shorter and more likely to answers

the user's information need. The primary motive to generate index able documents and to plan the unstructured corpus into some organised structure that would able to satisfy the target analysis as per user requirements.

The notion of a document is not clearly defined in lifelogging because of the continuous nature of lifelog data. Efforts have been made to impose a unit of retrieval such as minute [10], activity [11] and event [1], which are all document-centric units (detailed descriptions will be provided in Chapter 3). A document in lifelogging would be a logically related temporal subset of consecutive minutes, including lifelog images and other data. This document serves as a retrievable unit. We have found that a lifelog document generated by the conventional way of event segmentation algorithm is not necessarily the best option, and we will explore alternative units of retrieval to segment lifelog data. This will be introduced in Chapter 3 and will be discussed in Chapter 5.

A conventional approach to automatically segment streaming data is similar to segmentation of video into shots and scenes. This requires structuring personal data into discrete units [4] which can be semantically enriched to form the basis of a lifelog retrieval system. Event segmentation of lifelog data has received research interest for about a decade now. Early researches on lifelog segmentation have taken a static time-based approach, which generates fixed and inflexible indexable units and are referred to as user life experience events. These events can vary in the range from a few minutes to a few hours and it was introduced by Doherty et al. in 2007 and 2008 [1, 12].

From the literature review given in Chapter 2 and research work by Chen, Li, and Bolanos et al. in 2011, 2013, and 2015 [13–15] have used similar approaches to Doherty et al. [1, 12] for the segmentation of lifelog data with a minor difference in volumes and type of data sources. We use the approach introduced by Doherty [1] as our baseline for building basic segmentation algorithm. Along with this baseline approach [1], we propose two lifelog segmentation approaches i.e. static segmentation based on indexing-time and dynamic segmentation based on

query-time. It is presumed that dynamic segmentation based query-time approach would better support the user in searches for specific content as per needs. This will be explained in Chapter 6. In case of static segmentation, we present several scenarios where pre-defined boundaries would not be optimal for user and where events are too long or too short. Therefore, we re-implement our baseline static segmentation based on indexing-time approach. This will be explained in Chapter 5. [1] that incorporates existing state-of-the-art techniques [15, 16] and propose a new dynamic segmentation approach based on query-time. This will be explained in Chapter 6. We then test and compare both segmentation approaches using the publicly available LSC2018 lifelog dataset³ to identify the better segmentation algorithm.

1.3 Research Motivation

Consider that you have a software on your computer that will allow you to retrieve your memories with your friends, family or colleagues at work. Imagine being able to tell any special life story or incident to your children and grandchildren. Lifelogging is the platform where this scenario is being made possible. There are several devices available in the market to automatically capture almost all of our physical movements, text messages, emails or voice calls, web browsing, locations, sentimental situations, eating habits, etc. So why not use all this personal archive to retrieve and store the meaningful information for longer time? Several comparative evaluation benchmarking exercises have been carried out over the past decade to introduce lifelog retrieval tasks for community. This will be introduced in Chapter 2. These evaluation campaigns were initially focused on text retrieval. Retrieval tasks breached out to other data sources such as visual data, e.g. videos/images and biometric data with time. We observe the need of data pre-processing and segmentation as fundamental steps to build such information retrieval engines for lifelogging. Therefore, we propose the concept of

³LSC2018 Dataset is available at: <http://lsc.dcu.ie>

basic units of retrieval for searching meaningful segments from streaming lifelog data. In this dissertation, we motivate doing deeper analysis to identify lifelog segmentation approaches based on basic units of retrieval that would be able to develop retrieval engines for lifelog application domain.

There has been a surge of visual capturing devices such as camera phones and digital video recorders in recent years. What makes the content created by such devices differ from lifelog content is that they produce burst content. A camera phone for example takes conventional photos in sequence of one or more throughout the day. These datasets are naturally segmented into events or points in time based on the gaps between time of human generated data capture. In case of lifelogging, the data streams are continuous and there is no logical point of segmentation. Consider an individual wearing a modern wearable camera that may capture thousands of images in a day. We believe that the data should be organised efficiently to be usable by the user. In this research work, we have explored existing technologies of data segmentation to process and organise the lifelog data collected by different wearable sensors.

The closest representation of a lifelog document from past research is called an event as mentioned before. The event is semantically related period from an individual's daily life activity on a particular day. Events can differ in their size. For example, a small event with few seconds such as locking up the door or a long event spanning over few hours such as working on a laptop for 8 to 12 hours where an individual is trying to meet a deadline to complete PhD dissertation. We explore how individuals in a lifelog search engine use these documents. Figure 1.2 shows different lifelog events typically called user activities performed in a day such as peeling carrots, driving, shopping,. and transition between these variable-sized events such as taking blood samples, having breakfast, commuting to work, working on laptop, etc.

To evaluate this research work in an open and comparative manner, the dissertation writer coordinated with lifelog event segmentation (LEST) task under



Figure 1.2: Example of different types of lifelog events and the transition of these variable-sized lifelog events in a day.

the umbrella of NTCIR-13 evaluation workshop. NTCIR is a collaborative evaluation workshop that addresses different challenging tasks in information retrieval. It will be explained in further detail in Chapter 2. It releases rich multimodal dataset and defines research agenda across a good deal of information retrieval related domains. LEST uses large lifelog test collections, explore segmentation methods and label multimodal lifelog data into a discrete set of retrievable documents [8]. Referring to the literature review on lifelogging in Chapter 2, we find some loopholes in selecting lifelog users, identifying target use-case of lifelog data, selecting devices used to collect lifelog data, observing privacy concerns of the owners while sharing publicly or third party organisations and selecting a basic unit of retrieval in segmenting data [1, 12, 15]. This motivated us to consider important challenges that need to be undertaken before building an improved version of lifelog segmentation algorithm. We explain these challenges in the following section.

1.4 Research Challenges

In this section, we will explain the basic challenges faced and considered being resolved before developing the lifelog segmentation algorithms:

- **Users:** The selection of lifeloggers is one of the biggest challenge in lifelogging because it is difficult to find someone who would share personal archive for the longer time. A voluntary participation is important to maintain the trust between researcher and lifelogger. Besides the selection of participants, we also need to decide the number of participants. Large amount of data needs to be collected from participants. Selection criteria should be observed to maintain equal ratio in gender, age, occupation, demographic, region, etc. We have resolved this challenge by doing our preliminary case studies, explained in Chapter 3 . We collected small-scale lifelog data from 25 users over the period of 2 days each. We also organised one interactive session with all participating lifeloggers to know about their experience before and after collecting lifelog data. We also utilised publicly available lifelog test collections wherever available. This will be explained in Chapter 5 and 6.
- **Target audience and use-cases:** It is important to know the target audience and different use-cases to whom the end product will be released for every unfamiliar task. If someone wants to build a smart-home application for security and wellness of old-age people, the target audience will be elderly who are over 80 years of age and their caretakers. Another example where someone wants to build an interactive system for athletes to record all their physical activities, injuries and dietary routine, the target audience will be the athletes, coaches and doctors. Rather than focusing on a specific use-case, we focused on daily life human activities for retrieval purposes. So we maintained a diversity of the participants in the target audience.

- **Devices:** To build optimised retrieval engine, we need to collect sufficient sources of lifelog data so that user can retrieve life experiences. Though even many wearable devices are available in market one needs to consider cost, durability, efficiency, battery power, and number of sensors, state-of-the-art technologies, etc. In our research, we used wearable camera devices such as autographer and narrative clip considering cost effectiveness to gather visual lifelog data, various smart phone applications to collect biometric data, semantic locations, food/drink or mood logs.
- **Privacy and ethical concerns:** Continuous emergence and advancements in sensing technologies enable us to collect and publish lifelog data with ease so much so that widespread adaptation of lifelog technologies have potential to pose challenges for ensuring personal privacy of subjects and bystanders in lifelog data. It is the prime responsibility of a researcher to respect the sentiments of participating lifeloggers and their surroundings. We need to completely remove or partially hide the identifiable content available in the lifelog data as per user's intentions before making the data publicly available for research or any other purposes. We organised one user experiment to explain considerations on user privacy and their surrounding environment. This experiment is discussed in detail in Chapter 3.
- **Indexable units:** Usually multi-sensor lifelog data is bursty, so it needs to be organised in some indexable discrete units called lifelog documents. For deeper analysis, multi-sensor lifelog data need to be fused, aligned and synchronised. We will address this challenge in Chapter 3 and 5 by defining generating lifelog documents.
- **Unit of retrieval:** From past research work in lifelogging, authors have built retrieval engines based on minutes. We found that retrieval of specific information needs a more constructive and meaningful retrieval engine, which sometimes is difficult to retrieve in a minute-based system. Hence, we

developed an ontology to define units of retrieval in lifelogging and thus decided on the most appropriate unit of retrieval for this study. This work is detailed in Chapter 3.

- **User information needs:** Usually we need topic specific user queries to build an ad hoc lifelog retrieval engine. Publicly available lifelog test collections include several user requirements and were specifically designed to support a minute-based retrieval. In this research, we needed to develop new user query topics that reflect a wide range of query-types that would be expected to be used with lifelog collections and reflect conventional retrieval needs of lifelogger based on the proposals in [17]. This motivated us to develop different queries for lifelog retrieval systems in this work. We will address this challenge in Chapter 6 while discussing how we developed different query topics as per user requirements.

The major focus of this research is to improve conventional segmentation algorithms to segment continuous lifelog data by considering state-of-the-art segmentation approaches and identify the best approach. By following this consideration, we propose our research hypothesis and research questions in the following section.

1.5 Hypothesis and Research Questions

To address the previously mentioned challenges in lifelogging research, we propose one main hypothesis based on the research gap:

Hypothesis: *Dynamic (query-time) segmentation models for lifelog outperform conventional static (indexing-time) segmentation modals for ad hoc search through lifelog datasets.*

Associated with this hypothesis, we identify several research questions that we need to address to prove or disprove this hypothesis. The research questions are following:

RQ1. What are the basic privacy concerns that researchers must address when gathering and analysing passively captured visual lifelog data, and how researchers should use privacy-aware lifelog test collection to evaluate different approaches to lifelog data segmentation and retrieval?

Collecting private visual lifelog data over a longer time and publishing it to a third party organisation could be problematic without knowing the privacy concerns of the lifelogger who collected the data. We considered it as a first step in any ethical research into lifelogging domain before proceeding with more scientific research questions that showcase the technical progress of this research. Additionally, we identified the need of observing variance in privacy concerns and what amount of data should be available for experimentation by the research community. To address user privacy concerns, we conducted user study to capture lifelogger’s concerns when sharing the lifelog data across three specified categories i.e. *Only for Me*, *Family/Friends* and *Public*). We infer that the researchers need to understand the expectations of privacy of participating individual before publishing the lifelog data. We present a user study in Chapter 3 which will focus on important privacy considerations that influence how the privacy-aware lifelog test collections (NTCIR lifelog test collections) were published. This also addresses RQ2.

RQ2. How can we construct privacy-aware lifelog test collections to support comparative benchmarking for ad hoc retrieval tasks?

Lifelogging with rich personal data imposes challenging problems in multimodal lifelog retrieval and at the time we stated this research, there were no privacy-aware lifelog test collection available to test both event segmentation and lifelog retrieval. This leads to the second research question: how one can construct privacy-aware lifelog test collection to support comparative evaluation of many approaches to lifelog event segmentation and

ad hoc retrieval. This explains releasing privacy-aware lifelog test collection and introduce relevance judgement topics to ensure consideration of queries and relevance judgements for the ad hoc retrieval tasks in Chapter 3. The dissertation writer co-organised various evaluation workshops such as NTCIR [8, 10], ImageCLEF [18], and LSC2018 [19] for detailed understanding of releasing lifelog test collections to the public or third party organisations and motivating comparative evaluation (see appendixA.4).

RQ3. What are the various units of retrieval documents for ad hoc search through multimodal lifelog data?

This research question considers what exactly is a document for ad hoc lifelog retrieval and how these documents can be generated from continuous lifelog data streams. Multimedia information retrieval approaches can only be implemented over lifelog data when there are indexable documents. This requires fusion of many temporally aligned sources of data together and fusion of a time-period of sequential data which builds a document. We conclude that this is not a simple research question since there are many approaches that one can take which can either generate indexable documents at indexing time or form query-specific retrievable units in some post-query process. To answer this research question, we developed an ontology to describe various units of retrieval in lifelogging that is based on the 5 R's of memory access such as recalling, recollecting, reminiscing, reflecting and remembering intentions. This was introduced by Sellen and Whittaker in 2010 [17] and is explained in Chapter 3. We investigated various units of retrieval and found an appropriate one for this research work through experimentation as discussed in Chapter 5.

RQ4. What is the state-of-the-art indexing time segmentation approach?

It is our conjecture that a reasonable unit of information that is neither too short nor too long is required for lifelog data. Such a unit should map to life

experiences or information needs of the individual such as moments when a user is drinking coffee, talking with friends and eating food in a restaurant. This is a concept of an event that the diversity and semantic gap between lifelog images gives rise to the difficulty of accurately detecting event boundaries in daily life analogous to real-world user experiences. We expect that we can improve accuracy of segmentation algorithm from previous research questions RQ1, RQ2 and RQ3 by including additional multi-sensor lifelog data. However, a systematic approach is required to test the benefit of different sources of data, which leads us to our fourth research question. Therefore, we address and answer this research question in Chapter 5.

RQ5. Does a dynamic (query-time) document segmentation approach outperforms a state-of-the-art (indexing-time) lifelog segmentation approach for ad hoc lifelog retrieval?

An intrinsic motivation for dynamic query-time based segmentation of multi-sensor lifelog data is to assist lifeloggers to access information from their lifelogs. The primary focus of this research is to compare the results from dynamic query-time based segmentation approach against conventional static indexing-time based segmentation approach which leads to our last research question. This research question helps the community to develop an effective ad hoc lifelog retrieval engine for retrieving past life experience based on proposed segmentation models. In result, we will propose different static and dynamic lifelog segmentation approaches in Chapter 5 and 6 and compare the accuracy for both approaches.

To conclude this section, we have introduced the hypothesis associated with five research questions for this research work. These five major research questions help us organise an overall plan for this dissertation. To address these research questions, we developed various lifelog event segmentation and retrieval algorithms. We used several data sources and found the best approach to segment multimodal lifelog

data. We discuss the major contributions of this dissertation in next section.

1.6 Contributions of this Dissertation

First, the aim of this research is to present an up-to-date review of the literature and the state-of-the-art in conventional event segmentation domain of lifelog data. Next, by experimentally investigating and attaining suitable solutions to the research questions, the hypothesis will ideally be shown to hold true and thus, following contributions of this dissertation are proposed.

- [1] Developing and releasing privacy-aware lifelog test collections for the continuous comparative evaluation in retrieval tasks (RQ1 and RQ2).
- [2] Developing an ontology for units of retrieval in lifelogging to define lifelog documents (RQ3).
- [3] Developing various event segmentation approaches (at indexing time) by integrating visual features extracted from lifelog images and various data analytics techniques to detect event boundaries accurately (RQ4).
- [4] Investigating and developing query-time segmentation algorithms to provide the flexible segmentation mechanism to the user as per information needs (RQ5).

1.7 Production List

- [1] **Rashmi Gupta**, Cathal Gurrin (2019). Activity Recognition and Segmentation Approaches to Multimodal Lifelog Data. In: International Conference on Content-Based Multimedia Indexing (CBMI 2019), Sept 4-6, Dublin, Ireland.
- [2] **Rashmi Gupta**, Cathal Gurrin (2019). Dynamic Segmentation and Retrieval of Daily Life Activities from Multimodal Lifelogs. In Press: Perner P. (eds)

Machine Learning and Data Mining in Pattern Recognition (MLDM 2019).
Lecture Notes in Computer Science. Springer International Publishing.

- [3] **Rashmi Gupta**, Cathal Gurrin (2019). Manual Annotations in Dynamic Segmentation of Multimodal Lifelog Data. In: 3rd International Workshop on Annotation of useR Data for Ubiquitous Systems (ARDUOUS 2019), March 11-15, Kyoto, Japan.
- [4] **Rashmi Gupta**, Martin Crane, Cathal Gurrin (2019). Considerations on Privacy in the Era of Digitally Logged Lives. In: Online Information Review (OIR), December 2019.
- [5] **Rashmi Gupta**, Cathal Gurrin (2018). Approaches for event segmentation of visual lifelog data. In: 24th International Conference on Multimedia Modeling (MMM 2018), 5-7 Feb, 2018, Bangkok, Thailand.
- [6] **Rashmi Gupta** (2018) Considering documents in lifelog information retrieval. In: International Conference on Multimedia Retrieval ACM (ICMR 2018), 11-14 June 2018, Yokohama, Japan. ISBN 978-1-4503-5046-4.
- [7] Aaron Duane, **Rashmi Gupta**, Liting Zhou, Cathal Gurrin (2016). Visual Insights from Personal Lifelogs. In: Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, June 7-10, 2016 Tokyo Japan.
- [8] Duc Tien Dang Nguyen, Liting Zhou, **Rashmi Gupta**, Michael Riegler, Cathal Gurrin (2017) Building a disclosed lifelog dataset : challenges, principles and processes. In: Content-based Multimedia Indexing (CBMI 2017), 19-21 Jun 2017, Florence, Italy. ISBN 978-1-4503-5333-5 (In Press).
- [9] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, **Rashmi Gupta**, Rami Albatal, Duc-Tien Dang-Nguyen (2017) Overview of NTCIR-13 Lifelog-2 Task. In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies, December 5-8, 2017, Tokyo, Japan.

1.7.1 List of Poster Presentations

- [1] **Rashmi Gupta**, Cathal Gurrin (2019). Manual Annotations in Dynamic Segmentation of Multimodal Lifelog Data. In: 3rd International Workshop on Annotation of useR Data for Ubiquitous Systems (ARDUOUS), March 11-15, 2019 Kyoto, Japan.
- [2] **Rashmi Gupta**, Cathal Gurrin (2019). Manual Annotations in Dynamic Segmentation of Multimodal Lifelog Data. *Research Faculty Day*, DCU, Dublin, 2019.
- [3] **Rashmi Gupta**, Liting Zhou, Aaron Duane and Cathal Gurrin (2017). Exploring Lifelogs via Object-based querying. *Insight Augmented Human Demonstrator Showcase*, Croke Park, Dublin, 2017.
- [4] **Rashmi Gupta**, Cathal Gurrin, Liting Zhou and Duc-Tien Dang-Nguyen (2017). NTCIR 13 - Evaluation of Lifelog Search. *Research Faculty Day*, DCU, Dublin, 2017.
- [5] **Rashmi Gupta**. Approaches for Event Segmentation (2017). *Insight Student Conference*, UCC, Cork, 2017.
- [6] Aaron Duane, **Rashmi Gupta**, Liting Zhou and Cathal Gurrin (2016). Visual Insights from Personal Lifelogs: Insight at the NTCIR-12 Lifelog LIT Task.

1.8 Dissertation Outline

The chapters in this dissertation are summarised in Figure 1.3 and organised as follows:

Chapter 1 introduces an overview of this dissertation which includes brief introduction of lifelogging, research motivation, research challenges that need to be resolved, hypothesis and research questions, contributions to this study, list of productions and structure of this dissertation.

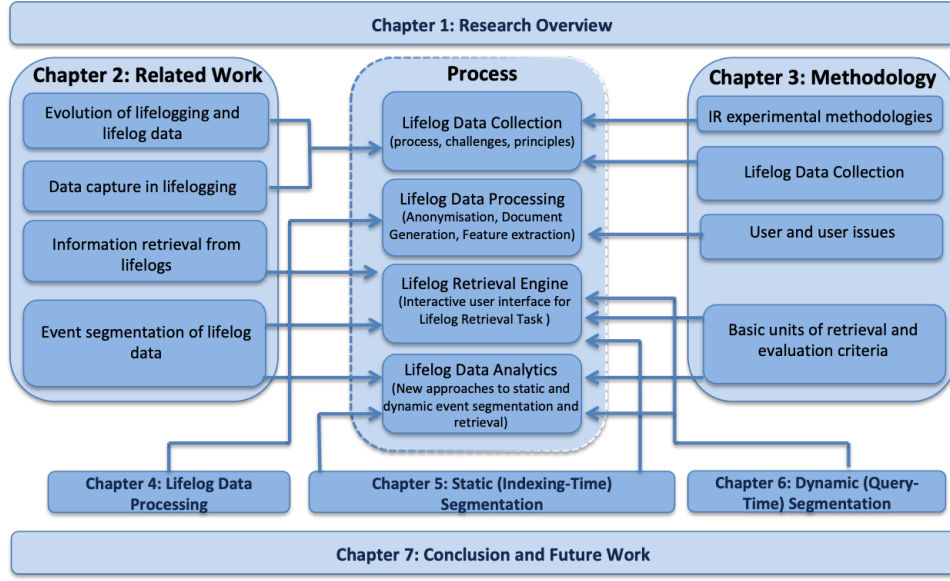


Figure 1.3: The Structure of this Dissertation.

Chapter 2 provides a comprehensive summary of related work and a critical analysis of existing event segmentation approaches. The first part of the chapter reviews the evolution of lifelogging and various lifelog devices available for collecting lifelog data. In next part, we present the previous research work on information retrieval from lifelogs and report the ongoing evaluation campaigns that organise the retrieval tasks for research community to test lifelog test collections comparatively. This chapter describes the existing indexing-time and query-time event segmentation techniques. This chapter concludes with a comprehensive review of lifelog segmentation approaches.

Chapter 3 presents the methodology used for this research work. In this chapter, we present the need of lifelog users and their issues such as selection, user interrogation and data governance. Lifelog data anonymisation section presents the prior work to release lifelog test collections publicly or to third party organisations which address our first research question, i.e. RQ1. In this section, one user experiment will be discussed that is focused on privacy concerns for lifeloggers, reason for participation in user study, introducing three different privacy levels i.e. private, family/friends and public and experiences of the

participants after experiment completion. Following that, this chapter discusses the process of releasing privacy-aware lifelog test collections to address our second research question, i.e. RQ2. We will then define basic units of retrieval and will identify the most appropriate unit for this research to address our third research question, i.e. RQ3.

Chapter 4 introduces the lifelog data processing techniques such as lifelog data cleaning, data organisation, feature extraction, and concept detection tools. This chapter then discusses the data analytic techniques used in this research work.

Chapter 5 presents the preliminary work and the baseline method used for building conventional static indexing-time based segmentation models using visual lifelog data. This chapter proposes two new segmentation approaches based on feature extraction using Caffe framework and Microsoft cognitive services API. It introduces various data analytic techniques and evaluation approaches such as the post-processing boundary gap and sliding window approach for system evaluation. Through experimentation, we identified the need of considering units of retrieval in this dissertation. Consequently, we discuss generating lifelog documents. Later in this chapter, we implemented and tested the newly proposed segmentation approaches using the publicly available lifelog search challenge (LSC) dataset and identified the most appropriate unit of retrieval i.e. activity to validate the segmentation results for the research community. This chapter concludes with a new enhanced way of segmenting lifelog data at indexing time that set forth the best basic event segmentation approach for this study which will resolve our fourth research question i.e. RQ4.

Chapter 6 presents the dynamic query-time based segmentation of multimodal lifelog data. This chapter focuses on activity based on dynamic query-time lifelog segmentation, and we improve the accuracy of the segmentation algorithm by considering manual annotations. The last experiment would compare both static indexing-time and dynamic query-time based segmentation approaches to identify the best approach for this research work. This chapter addresses our

fifth research question i.e. RQ5 and completes the decision to prove if the hypothesis is true or false.

Chapter 7 concludes this dissertation by analysing the current work and summarises the research methodology, re-iterating and discussing the solutions to the research questions, contribution and limitation of this study and future directions of this research based on the proposed methods.

Chapter 2

Related Work

2.1 Chapter Overview

This chapter presents a comprehensive literature review that mainly focuses on the critical analysis of existing approaches for event segmentation of multi-modal lifelog data. In section 2.2, we outlay the evolution of lifelogging. Various lifelogging tools for multi-sensor lifelog data collection is discussed in section 2.3. We also discuss various lifelog data sources and data pre-processing techniques in section 2.3. A detailed review of information retrieval methods is provided in section 2.4 and evaluation is given in section 2.5. We provide an up-to-date review of the existing static (indexing-time) lifelog segmentation approaches in section 2.6. We start with evolution of lifelogging in the following section.

2.2 Evolution of Lifelogging

Throughout the historical record of human’s life, people have tried to record daily life activities on paper diaries for recalling it at a later stage. However, this manually recording is time consuming and inefficient. For this work, we are focusing on paperless digital lifelogs. Digital lifelogging is a capturing of totality of an individual’s life activities. It provides a flexible mechanism to the user to access and analyse life experiences to observe sentimental or behavioural changes over time [20, 21], personal health and wellness [22, 23], change in dietary habits [24–26] and improve quality of life awareness with lifelogging [27]. Bush et al. in 1945

proposed Memex, a desk-based device which suggested a hypermedia system based on the user's life experiences [6]. Bush described the Memex as an “enlarged intimate supplement to one's memory”: enlarged (to gather and storing user lifelog data as much as possible), intimate (data should be private to the owner) and supplemental (to work in collaborative teams with one's memory). Smith et al. [28] in 2011 presented a definition of lifelogging that is “in essence, the collection of data to illustrate a person's life”, where they focused on building digital tools for data collection, storage, analysis, and display. Later in 2013, They described lifelog as the “diary to end all diaries — a multimedia, digital record of everywhere you go and everything you see, hear, read, say and touch”¹.

Over the past two decades, many more researchers started building lifelog projects inspired by Memex. For example, the information processing techniques office of the Defence Advanced Research Projects Agency (DARPA)² (at the U.S department of defence). They described the lifelog as a system that is “able to trace the 'threads' of an individual's life in terms of events, states, and relationships”. The system can store an individual's life experiences e.g. phone numbers dialled, email messages viewed, every breath taken, steps made and visited places. The primary goal of the DARPA Lifelog project was to build a predictive function that could predict the meaningful patterns in the continuous stream of temporal lifelog data. It could also observe the user's daily life routine, habits, and relationships with other people, organisations, locations, and objects. Gordon Bell [29] developed a lifetime digital data storage system, called MyLifeBits, as described in Chapter 1. It was based on the hypothetical Memex computer system described earlier. This system provided additional facility to record text and audio annotations, hyperlinks to the web content, reviewed documents, automatically captured pictures and sounds that an individual has

¹“Your life at your fingertips — courtesy of the Pentagon”. USA Today. 2 June 2003. Retrieved 20 December 2013.

²DARPA Lifelog System: <https://news.bbc.co.uk/2/hi/technology/3497596.stm> and <http://www.forum.nokia.com/main/0016-207000.html?model=6620> introduced Nokia Lifelog project.

experienced in a lifetime. Recently, researchers are being focused on building domain-specific lifelog systems. For example, Park et al. in 2010 [30] proposed a mobile based social network recommender system that used the user's lifelog data to send messages over phone such as "Your friend is nearby. Why don't you make a phone call to him?" to the user.

Yoshihara et al. in 2013 [31] introduced a health-care system for elderly people. This system collects lifelog data to analyse the daily routine and habits of the people and share this information with their caretakers and family members. Gelogo and Kim proposed a similar work in health-care monitoring systems [32] composed of wearable devices, a smartphone application and a medical record storage system. Such domain specific lifelog systems facilitate users to access lifelog data as per their needs. However, among the growing number of lifelog storage systems as reviewed above, the MyLifeBits lifelog system is one of the better known in the research community [33]. There has been a fast development of digital technology in lifelogging over the past decade. Therefore, we have investigated and reported the latest literature review on the evolution of digital technology in lifelogging in the following section.

2.3 Data Capture in Lifelogging

The research studies in lifelogging are being done with the help of one or more lifelog devices. The definition of lifelog device is extending from custom wearable devices to smartphone applications. The lifelogging wearable technology may incorporate special sensors such as accelerometer, thermometers, heart rate tracker, sleep tracker and/or GPS tracker in one or many embedded devices. One earliest customised lifelogging tool developed by Steve Mann in 1997, [34], called EyeTap Digital Eye Glass that gathered lifelog data such as images, audio and videos. We discuss the evolution of such lifelog devices and the use of these devices to collect lifelog data. It provides a holistic overview of captured lifelog data which would address one of

the key challenges of selection of devices as discussed in Chapter 1.

2.3.1 Raw Data Sources

- **Visual data sources:** It is our belief that the visual lifelog data which includes both images and videos captured from wearable devices are the most important source of data for quantified self-analysis. We present an up-to-date review on the advancements made in wearable cameras used over the past ten years. SenseCam is a wearable camera that takes the pictures automatically at approximately up to 2000 VGA images per day. The device has fisheye lens and sensors such as accelerometer, digital light sensor, heat sensing, audio recorder and passive infrared sensor for detecting any human presence. The researchers have used this device in many studies for instance, Karim et al. in 2006 [35] used SenseCam device to facilitate visually impaired people to decide during business meetings. Doherty in 2007 and 2008 by [1, 5] collected visual lifelog data using this wearable device. They proposed the lifelog segmentation algorithm based on a different frequency of each sensor data collected by SenseCam. Byrne et al. in 2009 [36] used the SenseCam device to collect lifelog data from nine participants to show limitations of capturing a large volume of multimodal lifelog data. They proposed the solution by collecting lifelogs with reduced content that could better reflect lifelogger's life events. Mair et al. in 2018 [37] used SenseCam images to support memory aids to both older and younger adults and showed age-related deficit for novel autobiographical material.

Autographer is another popular wearable device developed by a British company, the Oxford Metrics Group (OMG) Life in 2013. Many early researchers used this wearable device to analyse their lifelog data. For example, Chowdhury et al. in 2016 [38] developed an effective visualisation system by collecting lifelog data from 22 lifeloggers. Using this wearable device, researchers improved the quality of capturing user experiences and

making lifelogging tools more engaging. Signal et al. in 2017 [39] exposed increment in global prevalence of overweight and obesity in children. In this study, researchers made 168 children wear the wearable camera over the period of four consecutive days to examine frequency and nature of everyday exposure to food marketing.

Narrative clip is another wearable camera that has gained popularity among research community. This device is being used in many studies, for example, Fan et al. in 2018 [40]. The researchers collected visual lifelog data using narrative clip device from two participants over the period of five months. The principle motive was to provide a digital diary which can automatically label and summarise lifelog data. Jiang et al. in 2019 [41] designed an emotion-driven lifelog system by using this feature of the narrative clip. They proved that the system is responsive, efficient and user friendly while collecting mental activities for example emotions.

During the early 2010s, action cameras (GoPro)³ were used to facilitate athletes and extreme sports enthusiasts in creating a first-person view that captures oneself while driving, skating, surfing, rock climbing, etc. Molino et al. in 2017 [42] used this device to generate large amount of video data. Ragusa et al. in 2019 [43] proposed a location-based temporal segmentation algorithm for egocentric videos generated by chest-mounted GoPro device.

- **Self-Quantified data sources:** Self quantified i.e. biometric data is another source of information in lifelogging to target individual's health and wellness. Many biometric devices such as Fitbit fitness tracker, Beddit sleep monitor, Aura smart strap and Oura ring are being used to capture an individual's physical activities. Physical monitors such as heart rate, galvanic skin response, sleep duration, calories burn and intake, steps count, blood pressure, cholesterol level and body temperatures are recorded with the help

³<https://gopro.com/>

of these devices. We explain some of the most popular biometric devices being used by the research community:

- **FitBit:** Fitbit tracker is a good example of single sensor health monitoring lifelogging tool⁴. The tracker utilises an on-board accelerometer sensor to measure the data such as steps count, quality of sleep, calories burn count, heart rate, guided exercise, weather information through GPS connection. It also keeps track of any physical exercises such as running, cycling, swimming.
- **Aura smart strap:** Aura smart strap⁵ is good for analysing bio-impedance to observe the changes in the human body composition, activity tracking, and heart rate monitoring. The device helps to get the information about the active and reactive resistance of fat, bone and muscle issues. It also helps to record the size of each tissue and hydration level in the human body with the account of individual's physical characteristics such as weight, height, age and gender.
- **Beddit sleep monitor:** Beddit is a sleep tracking device to monitor sleep pattern of an individual. The device can record heart rate to show health, fitness and stress levels. It also can record breathing i.e. number of breaths per minute; snoring to show the poor sleep quality and environment condition such as room temperature and humidity. Brazendale et al. in 2019 [44] used this device to measure sleep patterns and total sleep time of school-aged children. The researchers then compared the sleep estimations from a field-based setting i.e. duration and bed/wake times at home across different measures in children.
- **Oura ring:** To keep the digital record of one's health and wellness, the sleep quality of an individual can provide beneficial insights. Oura ring was launched in 2018, to measure the impact of enough restorative sleep

⁴www.fitbit.com

⁵<https://auraband.io/>

on mental and physical health of an individual⁶. Dunn et al. in 2018 [45] used Oura ring to show the impact of using wearable sensors in health and clinical applications such as sleep, neurology, mental health and physical movement disorders. Later De Zambotti et al. in 2019 [46] used Oura ring and compared it with polysomnography. They found Oura ring had a 96% sensitivity to detect sleep, 65% in detecting ‘light sleep’, 51% in detecting ‘deep sleep’, and 61% REM sleep.

- **Data sources from smartphones and software:** There are several other lifelogging tools available in the market along with wearable devices. These tools facilitate the user to collect lifelog data in various forms such as location, mood and human-computer interaction data. One of the most popular mobile applications is moves app⁷. It provides a platform to record every location information visited by its user. To collect human computer interaction data in lifelogging, Hinbarji et al. in 2016 [47] built a comprehensive logging and visualisation tool for capturing computer usage activities. The activities such as keyboard activities to record key typed by user, mouse actions, clipboard actions to track all copy-paste operations, keystrokes dynamics to track time stamps of pressing keyboard’s buttons, application transitions i.e. an alternative method to track the current app, and screenshots of user’s current active window screen and its title associated with each screen shot. Another example of time tracking computer software is rescue time⁸. This software works in the background on computer, phone, or tablet and allows the individual to record and understand how efficiently the time is being spent on computer each day. The user can set duration of spending time on social networking sites and any other educational sites. User also gets the periodical alerts to meet goals, trends, weekly reports and offline time.

⁶<https://ouraring.com/>

⁷<https://moves-app.com/>

⁸<https://www.rescuetime.com/>

2.3.2 Descriptive Data Sources

Descriptive data sources provide broader understanding of even context, identity or object available in the lifelog content. Therrell et al. in 2019 [48] proposed a theory about collecting descriptive metadata as “More product, more process.” in digital image collections. This study emphasises the importance of metadata and revealed a lower level of descriptions for digital image data collection. A general definition of descriptive data defined by Dey et al. in 2001 [49] as “Descriptive data is any information that can characterise the situation of an entity. An entity is a person, place or object relevant for the interaction between the user and the application including the user and the applications”. Later, Dey [49] extended the definition as “Descriptive data is typically the information about location, identity and state of people, groups, and computational and physical objects”. We followed this definition and used following descriptive data to improve the analysis.

- **Date and time:** Date and time are used to identify the start and end point of user activities while implementing segmentation algorithms [1,12]. We believe that date and time provide a flexible retrieval mechanism to the user in case he/she wants to recall a specific lifelog content.
- **Manual log of personal behaviour:** The manual log of personal mood or social gatherings has been used to recognise the emotions of an individual. Gurrin et al. [8], maintained a manual log of mood such as happy, excited, sad in lifelog test collection was released in NTCIR 13 evaluation workshop. Cho et al. in 2018 [50] developed a human emotion recognition system that automatically recognise the human emotions. The system detected blood volume changes in the microvascular bed of tissue in the skin, GPS locations and ambient noise.
- **Dietary log:** We can observe several diseases because of unhealthy dietary routine of people. Researchers have captured the food and drink habits of an

individual over the past few years. The reason to analyse diet patterns was to provide the future indications of potential health hazards such as diabetes, obesity, diarrhoea or anaemia. Herruzo et al. in 2016 [51], developed an automatic diet recognition system by introducing 115 food classes with an average of 400 images per dish and 12 food categories with an average of 3800 images per class. The system recognises unhealthy food patterns of an individual and understands the individual's lifestyle.

- **Locations:** Most wearable devices have embedded GPS sensor which is also connected with a wireless system. GPS system provides three values to one's location such as latitude, longitude and altitude. These three dimensions of location are not understandable by the non-technical users. The transformation of these dimensions in some context could help user track their visited locations. This could provide easy access to lifelog content such as home, workplace, restaurant or church.
- **Human activity:** Wearable lifelog devices described earlier included accelerometer sensors to capture the individual's physical movements per each second. It is challenging to handle a huge number of numerical values that sensor captures even while sleeping. Assigning broader version of contexts such as running, walking, sitting and relaxing, gardening, standing still, bicycling, exercising, climbing stairs, indoor rock climbing, jumping, working on laptop, watching TV, etc. Many researchers [52–57] in previous study, used various machine learning algorithms such as fuzzy logic, decision tree, naïve bays algorithm, Markov models, k-nearest neighbour, and neural networks to detect human activity contextual data from raw accelerometer sensor data.
- **Environmental data:** Environmental data includes ambient temperature, light status, noise level and weather, etc. Primary consideration of this contextual data is to focus on “Where is the user and what he/she is doing”?

For example, the event can be easily detected in time where the user is in music concert or in church for praying based on the noise level or if the user is doing exercise or aerobics in a gym or doing meditation outside in a park on a sunny day. Doherty et al. in 2007 [12] proposed an event segmentation algorithm to segment lifelog data using temperature and environmental noise data.

- **Visual descriptive features:** Various feature extraction tools in pattern recognition, image processing or data visualisation have been used to build lifelog segmentation algorithms in the past decade. These tools extract features to retrieve informative and non-redundant derived values that lead to better human interventions by lifelog segmentation algorithm. In this section, we review low-quality and high-quality feature extraction tools that we have used:

- **MPEG-7:** MPEG-7 visual descriptor [58] generates low quality descriptors for colour, texture, shape and motion properties of the content. The commonly used colour descriptor includes dominant colour (DC), scalable colour (SC), colour structure (CST) and colour layout (CL). Texture descriptor includes homogeneous texture (HT), texture browsing and edge histogram (EH). Shape descriptor includes contour-based shape (CS) and region-based shape (RS). The regions for extraction can be from a complete image termed as global, a single region defined by a binary mask, all regions of a region map termed as segmentation mask and rectangular regions defined by their coordinates termed as Bounding boxes. An open source Visual Descriptor Extraction (VDE)⁹ application is available online for the extraction of the MPEG-7 visual descriptors. In information retrieval systems, we used these descriptors alone or in combined with other tools to make

⁹<http://image.ntua.gr/iva/files/uploads/tools/vde/documentation.pdf>

the relevance judgement with the descriptors of a query image.

- **Caffe framework:** Caffe is an Imagenet classification based framework [59] which has 1000 different classes for animals, person or any object. The Caffe framework [2] is CNN based deep learning framework which provides visual descriptive features such as pixel, motif, human or object, textual features such as character, word, clause, and sentence and speech features such as audio, band, phone or words.
- **Microsoft cognitive services API:** Microsoft research developed a deep learning framework called Microsoft cognitive toolkit (CNTK) [3]. CNTK toolkit is a set of APIs, SDKs and services available for developers and researchers. Using Microsoft azure machine learning APIs, the researchers can make more intelligent and user engaging applications. These applications can understand the user emotions and help in video detection, face detection, speech, text or vision recognition. For this dissertation, we used computer vision API to extract the high-quality visual features from lifelog images using CNTK toolkit.

In this research, we have used MPEG-7 visual descriptors to extract colour descriptors and edge histogram features from the visual lifelog images. We have also used high-quality visual descriptive features from Caffe framework and Microsoft cognitive services API such as image tags and categories; and descriptive metadata described earlier such as manual food and drink log, date and time, location, human activities and environmental data. More details of the lifelog data being used in this research will be presented in following chapters. We report the approaches of information retrieval from multimodal lifelog data in the following section.

2.4 Information Retrieval from Multimodal Lifelogs

The nature of retrieving information from lifelogs has changed in past two decades. Westermann et al. in 2006 [60] proposed an interactive retrieval system using audio, video and text data sources. They developed a high-level architecture of an e-chronicle that deployed steps such as data gathering, assimilation of gathered data to efficiently dealing with real-time data by keeping relevant content for the user retrieval and building interaction environment to facilitate the user access. Lee et al. in 2008 [61] collected lifelog data using SenseCam wearable device from one active lifelogger over the period of one year. They developed an online query-by-example style SenseCam image management system called “*My Visual Diary With SenseCam*”. User can browse for the lifelog content by selecting a day and visual summary with lifelog images using this system. Each image corresponded to a major event that happened during the day and provided one-glance view of that day. Events are refined and organised based on user annotations; events are listed in the user’s “favourites” and user account settings for the last presentation.

Kelly et al. in 2010 [62] introduced the importance of using biometric data as a source of query independent scoring in lifelog retrieval. They collected galvanic skin response and skin temperature data from three postgraduate students over the period of one month. They developed a retrieval engine using BM25 ranking model [63]. They also found an improvement in health-specific information retrieval by using skin temperature data. Later, Chennuru et al. in 2012 [64] developed a retrieval system using mobile data to record, index and understand the life of a mobile user. They divided the system in three phases such as lifelogger mobile client, lifelogger application server and lifelogger web interface. Lifelogger mobile client was used to wear a helmet with a Nokia N95 phone mounted and could capture accelerometer data, GPS locations, pictures, audio rotations and WiFi signal strength for indoor locations. Lifelogger application server was used to store, process and model the collected lifelog data in the form of user activities. Lifelogger web

interface was used to facilitate the user to browse, annotate and search their lifelogs. However, there is a similar approach introduced by Zhou et al. in 2013 [65], where the researchers developed a lifelog activity recording and management system called ZhiWo. It provides the user an activity recognition system for personal life help to use machine learning algorithm such as support vector machine (SVM).

Researchers are focused on retrieving domain-specific content from the lifelogs in recent past. We have explored the recent existing work introduced by Bahrainian et al. in 2016 [66], where they have developed a retrieval system that links social interactions of the user with corresponding images available in lifelog data. In addition, they developed a dynamic topic-modelling algorithm for indexing and searching personal memories of user. They collected social interaction data from one participant over 60 days such as audio clips, images using narrative clip, wearable device and GPS coordinates where the participant was part of social gatherings. We explored similar approaches over the past few years in 2017, 2018 and 2019 by Ionescu et al. [67–69] to retrieve meaningful information from lifelog test collections. The prime focus of lifelog and medical imaging task was to understand, retrieve and summarise daily life activities of users, caption analysis, predict the tuberculosis in images, medical visual question answering and segmenting/labelling lifelog images for 3D modelling.

Zhou et al. in 2018 [70] built as a baseline interactive lifelog retrieval system called LIFER. It used 24 user query topics introduced by lifelog search challenge. Though this was only a preliminary study of building a retrieval system, they identified some limitations such as improvement in query panel to provide the optimised query layout for user and embedding the self-query generated metadata to improve the performance of the system. Munzer et al. introduced a similar approach to the lifelog search and retrieval problem in 2018 [71, 72] where they developed a lifelog search engine called lifeXplore. They collected images and video lifelog data from one active lifelogger over the period of 27 days. This data is publicly available for the evaluation and known as Lifelog Search Challenge (LSC)

dataset.

2.4.1 IR Approach

Along with all information retrieval algorithms described above, we have explored the generic information retrieval approaches to rank model such as term weighting (TF-IDF) and best matching (BM-25) ranking algorithm that researchers are using to build lifelog retrieval engines these days.

- **Term weighting:** In information retrieval, TF-IDF [73] and Okapi BM25 [63] are two standard ranking algorithms. These ranking algorithms are used to rank best matching documents according to their relevance to a searched user query topic. We implemented both ranking algorithms and found little difference between the two. The Okapi BM25 outperforms the results of TF-IDF ranking technique.

TF refers to Term Frequency. It counts the frequency of term occurrence in the document based on its measure of importance.

IDF calculates the global document collection score that identifies the importance of the term to the document collection. It is inverse of a score called Document Frequency (DF) which is a count of the number of documents that a term occurs in.

TF*IDF weighting allows calculation of a term importance weight for the occurrence of a unique term in a document and is calculated using the following equation (2.1) below:

$$w_{ij} = tf_{ij} * \log\left[\frac{N}{df_i}\right] \quad (2.1)$$

where: w = the weight assigned to a term T in a document D , tf = frequency of term T in document D , N = number of documents in collection and d = number of documents where term T occurs at least once.

- **Okapi BM25** refers to Best Matching ranking algorithm used by search engines to rank matching documents according to their relevance to a user query topic. It is based on probabilistic bag-of-words retrieval function that ranks a set of documents based on the query terms appearing in each document regardless of the inter-relationship between the query terms within a document e.g. their relative proximity. It represents a whole family of scoring functions with slightly different components and parameters from state-of-the-art TF-IDF-like retrieval functions used in lifelog document retrieval where query Q , containing keywords $q_1, \dots, q_n$, the BM25 score of a document D is as (2.2) follows:

$$score(D, Q) = \sum_{i=1}^n IDF(q_i) * \left[\frac{f(q_i, D) * (k_1 + 1)}{f(q_i, D) + k_1 * (1 - b + b * \frac{|D|}{avgdl})} \right] \quad (2.2)$$

where $f(q, D)$ is q 's term frequency in the document D , $mod(D)$ is the length of the document D in words, and $avgdl$ is the average document length in the text collection from which we draw documents. k and b are free parameters, usually chosen in absence of an advanced optimisation. k belongs to $[1.2, 2.0]$ and $b = 0.75$. $IDF(q)$ is the IDF (inverse document frequency) weight of the query term q . It is usually computed as (2.3) below:

$$IDF(q_i) = \log \left[\frac{N - n(q_i) + 0.5}{n(q_i) + 0.5} \right] \quad (2.3)$$

where N is the total number of documents in the collection, and $n(q)$ is the number of documents containing q . In addition, we include ranking factors such as date, time and faceting; and facilitated operations such as logical operators, boolean operators, Ngrams and stemming to facilitate user queries. We will explain the implementation of these ranking algorithms in Chapter 6.

- **Learning-to-Rank:** More recent machine learning approaches such as

learning to rank for information retrieval (IR) tasks [74] automatically construct a ranking algorithm using training data. The algorithm can rank the documents according to their degrees of relevance, preference, or importance. Therefore, we can categorise the learning-to-rank algorithms into three approaches, such as pointwise, pairwise, and listwise approaches to analyse the query-level generalisation abilities.

Pointwise approach where the input space contains the feature vector of each single document and the output space contains the relevance degree of each single document.

Pairwise approach where the input space contains a pair of documents, both represented as feature vectors and the output space contains the pairwise preference (which takes values from 1, -1) between each pair of documents.

Listwise approach where the input space contains the entire group of documents associated with query q and the output space contains the relevance degrees of all the documents associated with that particular query. We consider this algorithm as the future direction and refer the readers to [74] for the detailed information about learning-to-rank techniques.

2.4.2 Evaluation of Lifelog Data Access

This section outlays the ongoing evaluation workshops for various information retrieval tasks in lifelogging domain. These workshops generate and publicly release the multimodal lifelog during the evaluation campaigns and highlight shortcomings of it. As a result, we report on the most popular evaluation benchmarking exercises such as TREC, NTCIR, LSC and ImageCLEF for lifelog information retrieval tasks in this section.

The Text REtrieval Conference (TREC) [75] is an ongoing series of evaluation workshops. It is mainly focused on a list of different information retrieval research domains. This workshop started in 1992 and is sponsored by National Institute of

Standard and Technology (NIST) and Intelligent Advanced Research Project Activity. This workshop popularised the test collection methodology and provided a large-scale evaluation infrastructure for the text retrieval challenges. It introduced the tasks e.g. relevance feedback task and spam task to propose spam filtering approaches. It also proposed web content tracking task to explore general information seeking behaviour of an individual while searching on a website. TRECVID known as video tracking task breached out to the videos that automatically segment, index and retrieve content-based information from digital video in the workshop. It also introduced the challenges with datasets and test problems regarding questions, user query topics or extracting target features. A uniform evaluation is being performed to provide participating lifeloggers a standard scoring mechanism to test their systems satisfactorily. The workshop organises a meeting or a conference for all the participants to gather and share novel ideas and present ongoing and future research directions in information retrieval tasks after system evaluation.

Later in 1997, the NTCIR evaluation workshop was launched. It is a Japanese counterpart of TREC evaluation workshop. The National Institute of Informatics (NII, a Japanese research institute) in 1999 started building test collections for developing IR systems in NTCIR. The primary goal of this workshop is to motivate the research in ad hoc information retrieval and cross-lingual information retrieval. One of the key challenges of this workshop is to investigate effective methods to construct and release large-scale test collections for reusable comparative evaluation of different retrieval tasks.

In this dissertation, we focus on the lifelog search task introduced by NTCIR evaluation workshop where it introduces four sub tasks over the series of ongoing workshops. The first subtask is Lifelog Semantic Access Task (LSAT) [76] based on known-item search task in an interactive or automatic manner. Lifelog Semantic Access (LSAT) Task introduces 24 different user query topics specifically focused on precision and recall [77]. The precision based user query topics reflect known or

remembered-item search tasks with one or a few relevant moments. The precision based evaluation identifies position of the first relevant moment for a user query topic. For example, query ‘when user was eating ice-cream beside the sea’ or find ‘when user was taking a photo of a bridge’. Recall based user query topics reflect information needs that can have many relevant moments. For example, find all instances ‘when a user was interacting with a dog’, ‘watching football’, or ‘crossing a bridge’. Each participated lifelogger provide a ranked list of results specific to their query and test in terms of mAP (mean average precision).

Second, Lifelog Insight Task (LIT) [76] concerned with knowledge mining from lifelogs with a focus on diet and health. Third, Lifelog Event Segmentation Task (LEST) [8] aimed to perform segmentation of multimodal lifelog data into specified human activities. This task introduced 16 human activity based query topics e.g. commuting to work, relaxing, preparing meals, using digital screens, etc. for the lifelog segmentation task. The participants provide a ranked list of user activities and evaluate the lifelog segmentation system in terms of precision, recall and F1-scores.

NTCIR-14 recently introduced one additional sub task called Lifelog Activity Detection Task (LADT) [76] to develop novel approaches to annotate multimodal lifelog data into 16 daily living human activities. The manual activity segmentation for over 45 days of lifelog data is achieved by segmenting data into fifteen distinct user activities e.g. travelling, face-to-face interaction, using computer, house working or any physical activity. It specifically focuses this task on precision scores, and participants provide relevant lifelog documents in terms of user activities. We will discuss the detailed information regarding user query topics and associated relevance judgements in Chapter 6.

A similar evaluation campaign was launched in 2000 from the European counterpart of TREC called ImageCLEF (Conference and Labs of the Evaluation Forum)¹⁰. This evaluation campaign ran for 3 years and promoted evaluation of

¹⁰<https://www.imageclef.org/2008/vcdt>

lifelog tasks for annotations, indexing and retrieval systems. The researchers proposed various approaches to build interactive lifelog search systems by participating at imageCLEF workshop. Another information retrieval evaluation workshop called lifelog search challenge (LSC2018) was launched in 2018 [78]. It introduced a wide range of 24 new user query topics to support interactive lifelog retrieval for the publicly available LSC2018 dataset. The authors developed 12 broad-focus queries e.g. find ‘where user is in the meeting’ or ‘watching tv at home’ and 12 narrow-focus topics to reflect the conventional retrieval needs e.g. find ‘where the user is talking with a person who has ponytail’ or find ‘where the user is waiting for a train at train station’. We summarise the above lifelog tasks at NTCIR evaluation workshop in Table 2.1 below:

Table 2.1: The summary details of query topics associated with specific lifelog tasks for publicly available lifelog test collections at NTCIR and LSC workshop.

Evaluation Campaign	Sub Task	Test Collection	No. of Days	No. of Topics
NTCIR-12/13/14	LSAT	NTCIR-12/13/14	90/45/45	24 (12 Precision + 12 Recall)
NTCIR-13	LEST	NTCIR-13	45	16
Lifelog Search Challenge	LSC	LSC2018	27	24 (12 Broad + 12 Narrow)
NTCIR-14	LADT	NTCIR-14	45	16

We have worked on three sub-tasks such as lifelog event segmentation (LEST), lifelog search challenge (LSC) and lifelog activity detection task (LADT) as described above while working on research leading to this dissertation. As we focused on identifying event and activity based lifelog segmentation models, we continued with lifelog event segmentation task (LEST) and lifelog activity detection task (LADT) task. The experimental explanation will be provided in Chapter 5 and 6.

2.5 Event Segmentation of Lifelog Data

Segmenting continuous streamed lifelog data into events as described in Chapter 1 has always been a key challenge for researchers. We present meaning of events and methods of segmenting lifelog data into events in this section. Zacks and Tversky

defined generic definition of an event as “an event is a segment of the time at a location that is conceived by an observer to have a beginning and an end” (as stated in Chapter 1). Westermann et al. [60] in 2006 described an event as “having three important dimensions: what it is (its name and class), where it took place and when it took place. A user navigating through an event database is interested in finding all events of a class that occurred at a particular location and time”. However, events could be a list of several sequential activities the user performed in a day such as “*I woke up at five, then went to the temple for meditation. Later, I prepared and had my breakfast; I started commuting to my work. In my office, I was working on the laptop for four hours and then I had lunch with my colleagues...*”. These statements are events. Each event has a description of start and end time stamps in continuous lifelog data. Automatic segmentation of lifelog data into events requires structuring personal data into discrete units [4]. We explore conventional lifelog segmentation approaches in the following section.

2.5.1 Approaches to Static Event Segmentation

We define static event segmentation as dividing lifelog data at indexing time where the lifelogger pre-define all user activity boundaries at the pre-processing phase for evaluation procedure. We will discuss it in the next Chapter. In NTCIR-13, the authors constructed a publicly available lifelog dataset that included visual capture, audio, biometric and information access data sources [79]. They coordinated event segmentation task [80] introduced by NTCIR-13 evaluation workshop. This was the first lifelog dataset that could support repeated and comparative evaluation of approaches to lifelog data event segmentation. This section presents various methods and approaches of event segmentation of multimodal lifelog data at indexing time.

- **Event segmentation based on text:** Textual data is the initial source data for event segmentation experiments. Stockes et al. in 2004 [81] proposed a course-grained natural language processing approach to segment broadcasted news stream using lexical chains. A topic segmentation tool was used to build

a set of tokens such as nouns, proper nouns, compound nouns in the news data to model discourse structure of the text. The relationship between tokens using WordNet thesaurus to detect specific news stories boundaries. Chakrabarti et al. in 2011 [82] proposed a model to segment and summarise the twitter feeds into events using HMMs. They calculated the cosine distance to measure the distance between two tweets and select those tweets which are closest to all other tweets from the event.

Allan [83] in 2012 utilised the event-based structure of broadcast news to find related news stories. The primary motivation was to segment the text into specific news stories and then monitor these stories for special unknown events. This system gathered the related stories or events into groups that associated with a single news topic. Story segmentation of content was the first step in TDT in which researchers [84] segmented a broadcast news into news stories using many potential sources of input such as audio, video, and text. They used several approaches such as finite state machine (FSM) with hidden markov model (HMM).

- **Event segmentation based on video:** Multimedia event detection began in applications such as video retrieval [85] where colour, motion and low-level visual features were used to segment video data into constituent scenes or shots. While there are many approaches to event segmentation or scene/shot detection from broadcast video, we will not examine these as they are outside the scope of this research. We will instead focus on wearable video cameras related to our work. Spriggs et al. in [86] proposed a method for simultaneous temporal segmentation and recognition of activity related to cooking and recipe preparation. The authors gathered video data at the same time from a single wearable video camera and multiple other static cameras, sensors, microphones. They used both sensor data along with visual GIST descriptors to describe the frames. The authors applied a

Gaussian mixture model for the unsupervised scene segmentation.

Bolanos [87] proposed an idea to segment the consecutive key frames of wearable video into three basic event classes. The event classes such as human movement: Static - where person or camera remains static, in transit - where person or camera is moving, and moving camera- where person remains still but camera moves. They used two approaches for segmenting videos into events. They extracted motion, colour and blurriness features from lifelog data and applied classifiers to get the rough estimation of the class labels. They applied energy-minimisation techniques based on graph cuts to identify spatial coherence of labels and separate consecutive frames in events. To improve the method implemented in [87], Poleg [88] proposed to classify the video frames into 12 different classes. The author based the classes on cues from wearer's head motion and implemented cumulative displacement curves that divide the image into a grid of $W \times H$ cells and compute the displacements. These cells refer to the motion in each image. Hence, integrating instantaneous displacements helped to detect long-term activity in video segmentation. Recently in 2018, Furnari et al. [89] introduced the temporal segmentation of egocentric videos for lifelogging applications based on 10 different personal locations visited by users and various negative ones, i.e. locations not specified by users. Thus, the proposed segmentation algorithm used by several applications related to lifelogging such as egocentric video indexing, retrieval and summarisation and estimation of amount of time spent at each specific location.

- **Event segmentation based on audio:** We believe that integrating audio data in our future work would provide useful insights to the research community. We can segment sound in an audio stream by examining the inter-similarity matrix of the Mel-Frequency Cepstrum Coefficients (MFCC) feature in the signal frames. Ghozi [90] implemented standard image

binarization based on automatic mean thresholding method. Median filtering was used to reduce the speckle-like resulting noise in the audio stream. We can easily identify the semantic scenes change from audio texture stream such as two different audio streams i.e. people crowd or clap sound from sports ground and birds chirping sound from the park.

Kucukbay [91] detected events in audio data using an adaptive features extraction scheme. They segmented the audio in 16 different audio events such as alert, clear throat, cough, door slam, drawer, keyboard, keys, knock, laughter, mouse, page turn, pen drop, phone, printer, speech and switch. They used MFCC and support vector machine (SVM) in recognising the audio events. Sadlier et al. in 2003 [92] proposed a combination of audio-visual contribution to event detection in field sports broadcast videos. They used speech-band energy tracking in audio and colour dominance pattern recognition in video. Jin et al. in 2004 [93] proposed an idea of automatic multi-speaker segmentation and clustering in meeting recordings. They recorded RT-04S meeting data with personal microphones for each participant and room microphones placed on the conference table. They used speaker diarisation error score to test the performance of the speaker segmentation.

- **Event segmentation based on continuous images:** Most closely related to the research in this report is the concept of event segmentation of lifelog photo streams. Bolanos et al. in 2015 [15] introduced a segmentation approach based on unsupervised hierarchical agglomerated clustering to extract the key-frame summary for the egocentric photo-streams. The authors tested the system over a dataset with 4005 images collected by 3 individuals and compared against a manual ground truth. Doherty in 2007 [12] implemented edge histogram low-level feature. They used Canny edge detection algorithm to capture the distribution of edges in lifelog images.

They also used low-quality features such as scalable colour i.e. white light levels, accelerometer values and temperature readings, but edge histogram proved a better representation of event boundaries. They implemented model tuning parameter such as Hearst's Text Tiling Algorithm [94] to determine the similarity between the adjacent block of images.

In line with an increase in volume of wearable camera data, Doherty in 2008 [1] introduced an event segmentation algorithm for wearable camera data. The authors extracted MPEG-7 visual features from images which lead to an improvement over previous approaches. They calculated various vector distance methods and showed that histogram intersection method and Euclidean distance method based on MPEG-7 features performed better than previous methods. In addition, they used automatic thresholding techniques such as kapur and mean thresholding as optimal thresholding techniques. Byrne et al. [16] presented an event segmentation technique based on content using five low-level MPEG-7 feature descriptors. They also used contextual information with light sensor i.e. changes in light and human motion sensor of the lifelog image dataset using bluetooth and GPS metadata. For event segmentation, they used heuristic split based on a long interval with no capture identifying the end of the event and sharp upward changes in frequency of capture identifying the start of the event.

Chen et al. in 2011 [13] gathered a large dataset of 450000 images of computer activities and 18 months of contextual data of 350000 records from 3 lifeloggers. They segmented the fusion of this rich lifelog data based on computer activities, location, and visual concepts using a Text-Tiling algorithm. Li et al. in 2013 [14] used event segmentation based on multi-sensor data recorded by a wearable camera with associated gyroscope and accelerometer sensors. They extracted the sum of all standard deviations (S-STD) feature and fine-tuned gyroscope data to enhance the performance to generate event boundaries. They utilised mean of HSV

histograms (Segment-HSV) feature in the study.

Yamamoto et al. [95] in 2017 took part in Lifelog Event Segmentation (LEST) sub-task of the NTCIR-13 lifelog-2 workshop [8] and proposed new segmentation approaches based on visual concepts and location data. They proposed similarity based segmentation approach to calculate the cosine similarity between the consecutive lifelog images. They used D-star based segmentation approach that focused on positions and time to extract the similarity neighbourhoods. Molino et al. [96] in 2018 proposed a new static event segmentation approach to predict upcoming next temporal segments based on past segments and trained over the lifelog datasets i.e. CLEF, NTCIR and EDUB-seg and R3. Gupta et al. [80] in 2018 proposed two new approaches for static event segmentation based on visual concepts from Caffe framework [2] and image categorisation by using concept detectors from Microsoft cognitive services API. They found image categories from Microsoft cognitive services API as best performing approach for event segmentation. This approach is based on pre-existing static approach [1] and is shown to outperform the existing state-of-the-art approaches discussed earlier in this section [16, 87, 89, 95, 96].

2.5.1.1 Conclusion to this Review

In this section, we reviewed the previously implemented experimental studies to segment lifelog data. Researchers have used different volumes of image lifelog data and sometimes fused visual data with other data from different sensors from several participants over a short period. The researchers extracted low-quality visual features and implemented clustering techniques such as hierarchical agglomerated clustering to segment daily life activities of the day into events. To conclude, we discuss different segmentation approaches, present a summary of lifelog data sources and associated content retrieved based on different scales or units such as full document, atomic unit, temporal unit and events in the lifelog

segmentation tasks (shown in Figure 2.2).

	WWW	Photo	Audio	Video	Twitter and Text	Lifelogs
Full Document	Web Page	image	sound	video (FD)	tweets	mm collection stream
Atomic Unit	image/video	-	-	keyframe	-	multiple au type
Temporal Unit	-	-	sub audio unit	video seg	-	1-5 minute units
Event	-	ES	audio ES	video events	multiple twitter docs	lifestyle events

Table 2.2: Summary of lifelog data sources and associated content retrieved based on different scales or units such as full document, atomic unit (au), temporal unit and event segmentation.

In this dissertation, we propose novel approaches of event segmentation by taking state-of-the-art approaches from past work [1]. We comparatively test proposed approaches and improve algorithms with additional sources of data and metadata. In addition, we develop new evaluation methods and datasets to support this comparative evaluation where our datasets are larger than the original datasets used in many of previously mentioned experiments. We also define the evaluation methodology that provides a fair and repeatable comparison between the different approaches.

2.6 Chapter Summary

In this chapter, we have reviewed the evolution of lifelogging, and digital technology used in lifelogging. We have also explored various evaluation campaigns launched till date and state-of-the-art approaches to segmentation of multimodal lifelog data. In addition, we reported the technical background in feature extraction tools and ranking algorithms used in information retrieval tasks. From literature, we have inferred that feature extraction from various lifelog data sources have proved to be the initial task while building lifelog segmentation algorithms. We identified and explored two different ways of segmenting multimodal lifelog data in this chapter. First, the conventional static (indexing-time) segmentation where we reviewed the segmentation approaches to all available data sources such as text, audio, video and images. Though our research focus is on visual lifelog data as the prime source of lifelog data, we

reviewed the second existing work in dynamic (query-time) segmentation approach i.e. IR from lifelogs to segment multimodal lifelog data. We will discuss the proposed research methodology for this dissertation in the next chapter.

Chapter 3

Research Methodology and Research Configuration

3.1 Chapter Overview

This chapter provides an overview of experimental methodology of the lifelog test collection which we followed in this dissertation. It supports comparative and repeatable experimentation to retrieve meaningful information from multi-sensor lifelog data as explained in Figure 3.1). In Chapter 1, we introduced the problem definition, research objectives and primary motivation to do this research. We proposed research hypothesis and research questions that need to be answered in this dissertation. In Chapter 2, we investigated the rationale behind study and review of existing state-of-the-art techniques associated with the research questions. In this chapter, we highlight the key challenges to collect lifelog data and introduce the methods used to collect it. It discusses the requirement of collecting lifelog data from multiple sources and useful data sources to build reusable lifelog test collections for several use-cases in lifelogging in section 3.3. Another challenge is to select the voluntary participant lifeloggers willing to collect lifelog data over the specified period of time for the study. We critically analyse privacy and data protection issues and present methods which we used to resolve these issues while selecting lifeloggers in section 3.4.

In section 3.5, we discuss a case study considering need of preserving anonymity of an individual in lifelog data while publishing data to the public and we will discuss

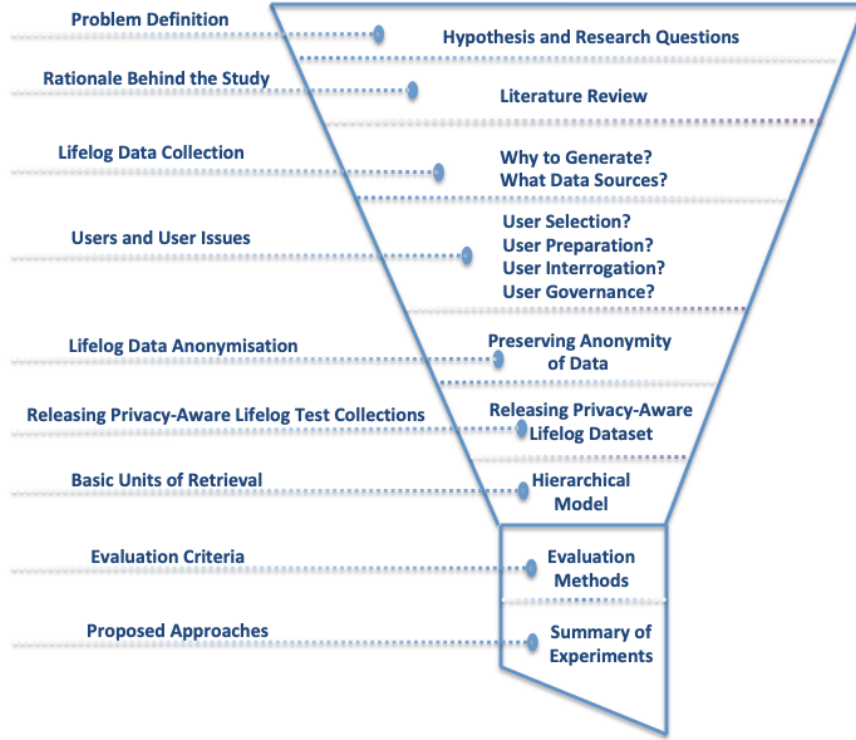


Figure 3.1: The experimental methodology for information retrieval in lifelog data.

how we use these findings in our research work. After anonymisation (see Figure 3.1), we define the process of releasing privacy-aware lifelog test collections publicly and discuss lifelog datasets being used in section 3.6. We propose a hierarchical model of unit for lifelog data retrieval. This helps to define lifelog documents into appropriate size indexable units as explained in section 3.7. Evaluation techniques have been briefly discussed in section 3.8. In addition, we report all the experiments conducted during the research using this methodology in section 3.9. The chapter is summarised in section 3.10.

3.2 Lifelog Data Collections

The fundamental challenges faced during collection of multi-sensor lifelog data are discussed in this chapter. It is our belief that a researcher cannot rely only on one or limited sources of data while developing ad hoc search engines for personal recall

of past life experiences. A large-scale corpus with limited sources of lifelog data e.g. visual lifelog images or semantic locations could lead to a very limited scope of target analysis. It should be noted that the dissertation writer was a co-organiser of the lifelog event segmentation task (LEST) at NTCIR evaluation workshop and she needed to provide her opinion on what sources of data would be needed and prove to be beneficial in LEST task for target analysis. The primary motivation of collecting personal data using various sensors by a number of individuals over a specified period of time is to build re-usable multi-sensor lifelog test collection. This test collection can be used for comparative evaluation and retrieval tasks which could benefit research community in developing query-agnostic information retrieval systems as per the users requirement.

Additionally, the research team was dealing with sensitive lifelog data the prime responsibility of dissertation writer (see appendix A.4) was to ensure the willingness of the users to share data publicly and with third party organisations for research. However, finding people who were willing to donate months or years of their personal data was a challenge. Privacy and data security have implications for both individual and society as a whole and needed to be considered. This raised the challenge of how to filter out the private information to protect the lifeloggers and individuals that appear in the lifelogs. To resolve this issue, the dissertation writer conducted a user study to understand the variance of privacy preference across participants while sharing their lifelog data publicly or to third party organisations. This user study helped the research team to understand user privacy concerns and which data governance guidelines need to be implemented. This will be explained in Section 3.4. This study provided insight into potential data governance guidelines to the research team. It also provided insights to the researchers/third-party organisations building digital trust and best practices for data sharing. Hence, the privacy-aware lifelog test collection that supports comparative and repeatable experimentation in this research was generated.

For this dissertation, we have collected and released both small and large lifelog

test collections. The challenges faced during dataset collection are introduced in this section. Process of releasing lifelog dataset will be explained later in this chapter. We explain the major challenges in building lifelog test collections as follows:

3.2.1 Challenges of Building Test Collection

- **What to log?** Following the spirit of the definition by Dodge and Kitchen [5], a lifelog typically comprises numerous different types of data captured from different sources. Ideally, we should log all information from all sources; however, it is not feasible in practice. Thus, deciding on what to log is a non-trivial task. We will explain how we have decided the rationale behind selecting different sources of data based on target analysis in following chapters.
- **How often to log?** Lifelog data shows considerable variance to capture velocity. Some sensors, such as biometric sensor, can capture data every second, whereas wearable cameras may capture between 1 to 5 images per minute. To be useful for the individual, we organise the lifelog content in minute long document and provide the retrieval facilities that could operate over data at different velocities and frequencies in order to address a wide variety of use-cases.
- **How willing are users to share data?** The success of the quantified-self movement [97] shows the willingness of individuals to gather archives of their life. People tend to be familiar with the concept of gathering their personal data or even being a subject in a lifelog of someone else. However, finding people who are willing to donate months or years of their personal data is a challenge. To resolve this, we did a user study to understand the variance of privacy preference across participants while sharing their lifelog data publicly or to third party organisations that helped us to generate privacy-aware lifelog test collection. This user study supported comparative and repeatable experimentation in this research. This will be explained in detail later in this

chapter.

- **What to share?** Privacy and data security have implications for both individual and society as a whole [98]. These are important issues that needed to be considered. Personal Identifiable Information (PII) such as personal ID numbers, car plates, addresses and others should not be shared. This raises the challenge of how to filter out such information in order to protect the lifeloggers as well as individuals that appear in the lifelogs. We have understood this challenge by conducting a user study that will be explained later in this chapter.
- **Who can access the data?** The goal of a publishing dataset is to share with researchers. Lifelogs, however, differ from more traditional shared datasets [99] (for instance the privacy-aware issues) and thus need a well-designed strategy to control data access and to change it if needed.

3.2.2 Lifelog Data Sources Used in this Research

Lifelog data collection involves the utilisation of multiple sources of unstructured, continuous and raw lifelog data from wearable, fixed, and informational e.g. software/online sensors that initiate the process of building a reusable lifelog test collection. Although there are multiply data sources which vary , we utilise the following data sources in this dissertation:

- **Visual information:** From the point of view of the wearer, the visual data was gathered using wearable camera such as autographer and/or narrative clip¹. These wearable devices take photo of landscape at a variable interval which is typically every 30 seconds and produce about 1500 to 2000 images per day. The images would include all the daily life activities of the wearer from breakfast to sleep. In lifelogging, visual lifelog data is the primary source of information to implement image processing and information retrieval tasks.

¹<http://getnarrative.com/>

- **Personal biometrics:** Using a smart watch² or equivalent wearable device, we can get all-day human biometrics such as heart rate, calorie burn and step count. Additionally, we can include daily blood pressure and blood sugar measurements. This biometric data could provide additional supportive data in understanding physical wellness of the individual that supports behavioural or sentimental analysis, when required, in lifelogging.
- **Human activity:** Logging an individual's physical activities can be done using a smart phone app e.g. Moves app³) that automatically records user activity in terms of semantic locations visited such as home, work or restaurant and physical activities such as walking, running and transport without requiring any user intervention. This type of content helps the community to observe the difference in user activities. For example, taking two regular user activities such as a person eating food and another person is socialising. In this scenario, one can distinguish these two regular activities based on the location data such as eating food could relate to having meal at any eating place such as home or workplace but socialising would relate to having meal in bar, restaurant or any other location where any social activities are happening.
- **Dataset metadata:** Associated with the visual images were various forms of metadata extracted from visual concept detector tools such as date, time; colour histograms, physical measurement such as age, gender, height, weight, music listening history such song name, album name, and artist name, and manual log of food and drink. We also appended additional data to the collection that is useful for retrieval, such as weather status e.g. rain, fog, sunny, light showers along with temperature conditions and day status i.e. early morning, morning, afternoon, evening, night, late night.

²<https://www.fitbit.com/>

³<https://moves-app.com/>

3.3 Users and User Issues

Before using lifelog data for experimentation, it is worth noticing that we need to put extra attention towards selection of participants who are going to participate in our experimental studies and their associated privacy concerns. Thus in this section, we discuss user participation, their issues and how we addressed their issues in this dissertation as follows:

3.3.1 User Selection

Lifelogging has a broad vision of building different domain-specific applications in research area using different types of data sources. As a result, it is challenging to decide the number of participants who are going to be engaged in this research and identifying the target usage of gathered lifelog data. While defining criteria for lifelogger selection, we observed that most of the people other than research community i.e. university students or professionals get very conscious towards their health and wellness. They are willing to capture their daily living records so they can observe their daily life activities including exercise routine, eating habits, sleep quality and long-term behaviour. It is our belief that the participants should be willing to log their own personal digital record and must be familiar with gathering lifelog data. They should also understand the presence of other people in their lifelogs. However, finding such people who are willing to devote months or years or even their whole life is a challenge. Consequently, for the user study employed in this dissertation, we targeted most of the participants in research or with academic background. These participants are believed to achieve the target analysis requirements. Detailed information about participants of this experimental study will be explained in following chapters.

3.3.2 User Preparation

One of the major responsibilities of the researcher conducting such experimental study is to make participants fully understand their research contribution. To achieve this, we have introduced an informed consent form to the participants that is the cornerstone of research ethics. The research ethics explain what our research is about, what the users participation in our project will entail and any risks that may be involved. Only after we have conveyed this information to the participants and once they fully understood the research in which their personal data is going to be used, we have got the permission to include them as voluntary participants. Thus, we considered the importance of making the participants properly satisfied to take their voluntary initiation to gather lifelog data.

3.3.3 User Interrogation

We focused on organising interactive user interviews at appropriate times to get mutual understanding with the user so we could change our research efforts as per user requirements. As part of the interview process, we conducted one user experiment with 25 users who collected lifelog data over the period of 2 days each. Understanding privacy concerns of participants in digitally logged lifelog data was the primary motivation behind this study.

3.3.4 Data Governance

Lifelog data governance is the overall management of the collection, integration, and protection of the lifelog data while being stored for research purposes. For management of lifelog data, we implemented different approaches such as cleaning raw unstructured lifelog data; preserving the anonymity of identifiable persons, objects, or brand names; synchronisation and alignment of multi-sensor lifelog data to organise it in indexable units and eventual deletion of the data; and include addition supportive contextual data and metadata without sacrificing

lifelog data quality. Lifelog data governance typically includes two main principles [100,101], discussed as follows:

3.3.4.1 Data Protection

Lifelog data is private to the wearer. There are privacy and data security implications for both wearer and society, including subjects and bystanders. These are important issues that need to be considered. Removal of any personal sensitive information such as debit/credit card details, passwords, home addresses, phone numbers, human faces, or any kind of human voice that can be easily identifiable is carried out to protect identity of individual in lifelog digital data. The data should be completely under the control of lifelogger who collected lifelog data and the researcher who has access to lifelogger's data . The data must also be completely anonymized so it would be difficult for the non-trusted entity to de-anonymize the actual content.

3.3.4.2 Data and Ethics

A conceptual challenge in exploring the role of ethical data in lifelogging motivates the interpersonal trust and relations among the lifeloggers, researchers, and third-party organisations. Moreover, gathering and storing lifelog data introduces the potential guidelines from research ethics that provide insights to the researchers/third-party organisations building digital trust and best practices for data sharing. For an instance, Kelly et al. [102] introduced an ethical framework that formalised the confidentiality controls over potentially unflattering and unwanted visual content using automated wearable cameras in health behaviour research. Consequently, they proposed four main ethical guidelines. First, informed written consent of participant where they recorded and ensured the count and duration of data collection, nature and type of collected data and removal of any content which contains nudity or unflattering body positions. Second, privacy and confidentiality where the researchers ensured the storage of collected data

according to national data protection regulations and the participants were allowed to remove the wearable device at any time. Third, non-maleficence where the participants instructed to remove the device in case of any intrusive argument against capturing data or the participants feel uneasy, threatened or uncomfortable wearing the device. Fourth, autonomy of third parties where participants need to seek the permission from their family members or cohabitants.

We have followed the above-mentioned ethical guidelines in this dissertation that helped us to collect and release privacy-aware lifelog test collection. We extend the previous work introduced in [102] by organising the post experience interviews with the participants to support the recollection and reminiscing of memory access [103]. Therefore, we now introduce the principles for lifelog data ethics as follows:

- **The highest priority is to preserve the anonymity of data:** Respecting the person behind the lifelog data and privacy concerns should be the primary concern before releasing it publicly or to the third party organisations. By implementing this principle, we encourage usage of informed written consent of participants and respect the autonomy of family members or cohabitants.
- **Downstream the usage of lifelog data:** Releasing lifelog data to different downstream levels such as private e.g. keep it for oneself or direct permanent deletion, semi-private e.g. shared with family/friends and public helps to make data consistent with the intentions and understanding of the disclosing party as explained in study in next section. We enforced this principle in this dissertation and will be explained in Chapter 5 and 6.
- **Secure storage and disposal of lifelog data:** Storing lifelog data at a trusted location that is both secure and easily accessible for further manipulations and ensuring removal of data from servers after particular period of time to prevent future release or use makes the lifelog data more ethical in practice. This principle motivates privacy and confidentiality ethical guideline introduced in [102].

We follow the principles for lifelog data ethics to address the lifelog data collection challenges such as willingness of participants to share data, content of share and who can access the data as discussed earlier. We present a user study to consider the anonymisation of lifelog data before releasing it to public. In the following section, we dive deep into the need of understanding the variance of privacy concerns across participants for their collected lifelog data that make us understand the process of generating privacy-aware lifelog test collections.

Lifelog Data Anonymisation In lifelogging, preserving the privacy of individuals inadvertently captured in lifelog data becomes an important consideration [4]. This concern need to be taken into account when sharing lifelog data on social media or when generating reusable lifelog datasets. Visual data from lifelog cameras captures not only lifelogger’s personal details but also identifiable or non-identifiable people around the lifelogger, and most times without their consent. Therefore, we need to hide the identity of people and any sensitive personal, professional or social information before publishing data publicly or to a third party organisation. To address the first research question (RQ1 described in Chapter 1, i.e. the expectation of privacy for individuals when gathering and sharing passive capture visual lifelog data and how researchers can utilise the privacy-aware lifelog test collections for evaluation of approaches to lifelog data segmentation and retrieval), we organised an interactive session with participating lifeloggers to gather their concerns and feedback about sharing lifelog data in our study. We now explain the experimental configuration and finding of this user study in following subsections.

3.3.5 Experimental Configuration

In this user study, 25 participants belonging to different backgrounds participated. This was done so we did not end up with a cohort of a single type of individual e.g. university students. We have categorised these participants into one of three broad types as summarised in Table ??; 10 researchers, 7 professionals and 8 others such as undergraduates or home-makers. There were 10 female participants and 15

males, all within the age range of 20 to 40. The three broad types of participant - researcher, professional and other are defined thus:

- **Researcher:** Researcher participants belong to an academic environment or professional research community. The experimenters observed that the researchers collected a comparatively large amount of personal data by wearing the camera all-day with few interruptions. Therefore, this group tended to have a higher proportion of data labelled as private (average 40%, see Table ??).
- **Professional:** Professional participants belonged to industry or non-academic organisations. They collected slightly less lifelog data than members of either of the other two categories and they are happy to share most of their lifelog data with family/friends (average 47%, see Table ??).
- **Other:** Other participants such as undergraduate university students or home-makers collect comparatively more data than professionals but less than the researchers. Generally they collected less private content in their lifelog data and were happy to share most of their data publicly (average 46%, see Table ??).

At the first stage, all participants were asked to wear an autographer wearable camera for 2 days each. The participants were then asked to answer a questionnaire which verified the rationale of them being a volunteer to collect lifelog data in this user study. The authors noted a diversity in the responses (explained in [103]). For example, participants agreed to participate because they found this study as a representation a novel and interesting activity. Some of the participants were prepared to collect lifelog data for self-observance and self-improvement in daily life activities and to capture memorable experiences for future recall etc. Perhaps due to the nature of this study which sought data during a short-time period, authors observed that the participants were not concerned to collect lifelog data from any

specific occasion or event which differs from previous findings [104]. It motivated us to collect lifelog data over a longer period of time to capture totality of an individual's life experiences which will be explained later in this Chapter.

At second stage, authors conducted a privacy assessment on the types of visual lifelog data that participants are willing to share by asking them to segment their data into overlapping categories with different access levels from personal to public. This work as done as motivated by the work in [104,105]). The three different access levels are: *Only for Me* (where the images are not to be shared with anyone but to be kept in a private lifelog archive); *Family and Friends* (where the images that will be shared only with family and friends and *Public* (where the images to be shared with anyone including family and friends. This data will be out of the control of the lifelogger once it is shared. So there is natural risk of data can be used to do harm to the lifeloggers or others. In addition, authors showed the variance of lifelog data (see anonymisation approach in Table ??) and authors allow the research community to gain the overall variance of privacy preferences for each category (i.e. 28% for Only for Me, 40% for Family/Friends, and 32% for Public) across the participants.

The main idea behind considering each image in one of these three categories/access levels is that such a categorisation should be easily understandable given that it is the typical privacy permissions of data in modern computing systems e.g. file or folder is being shared with owner, group, or everyone as discussed in [106] and [107]. It is worth noting that some data may not be suitable even for private storage, which could be dependent on the sensitivities of the lifelogger or other external/legal factors. For example, images captured accidentally in a bathroom or other private situations are likely not to be suitable for permanent storage even if it is kept private to the lifelogger. Such data is not used in this research because it was deleted by the participants in a pre-experiment data cleaning phase as described later. We refer [103] to the reader for the detailed discussion on the findings for this stage.

At third and final stage, after completion of user study, each lifelogger was

asked to share his/her personal experience as to how did the wearable device affected their lifestyle in both positive and negative manner. This was done to ensure the development of consent based privacy-aware lifelog test collections. This analysis revealed a number of experiences such as getting self-realisation and using it to improve their daily lifestyle in terms of food habits, exercise routine or working hours. Participants felt self-motivated and happy to review their lifestyle and self-reminded to turn off the camera in bathrooms or while engaged in any intimate moments. However, some participants who belong to research community represented as expert researchers reported that self-data gathering was a normal part of their daily life and they did not even notice wearing camera.

3.3.6 Results and Findings

In Table ??, authors showed variance of lifelog data across all three user types based on access level categories. For each user and user type, we could see the willingness to either share or keep lifelog data private. Authors observed that the researchers are willing to keep most of data private, professionals are happy to share their data with their family and friends and the others are happy to share much of their data to the public. In this section, we engaged in a deeper analysis of the findings presented in Table ?? and we discuss the privacy assessment by participants for their own visual lifelog data in detail and across all three experimental stages.

3.3.6.1 Stage 1: Reasons for participation in user study

To understand the reasons why participants had taken part in the study and their thoughts on visual lifelogging, each participant was asked the reason(s) why they voluntarily participated in this experiment and authors noted a diversity in the responses, as summarised in Table 3.2. 76% of the participants comprising of 9 researchers 4 professionals and 6 others agreed to participate because they found study to represent a novel and interesting activity. While 60% of the participants which included 10 researchers and 5 other participants were prepared to collect

lifelog data for self-observance and self-improvement in daily life activities. 52% of participants comprising 3 researchers, 6 professionals and 4 others took part to capture memorable experiences for future access. Perhaps due to the nature of our study which sought data during a short-time period, we observed that the participants were not concerned to collect lifelog data from any specific occasion or event which differ from previous findings [104].

Reasons for the Participation	Number of participants (N = 25) and Overall (%)
New and interesting thing to do	19 (76%)
Self observance and self improvement in daily routine	15 (60%)
To recollect the memories for future	13 (52%)
Capturing/sharing important moments with your family	6 (24%)
For being social and sharing personal activities with society	4 (16%)
Just for fun	2 (8%)
May helpful for future research	1 (4%)

Table 3.2: Stage 1: Summary results of reasons for being participant in user study.

3.3.6.2 Stage 2: Three different privacy levels

The authors analyse privacy assessment by participants for the different types of content captured in their lifelog data by exploring their willingness to share across all the three access levels i.e. *Private (Only for me)*; *Semi-Private (Family/ friends)*; and *Public*. Authors did this by examining the volume and type of lifelog data that participants categorised.

1. **Private:** Table 3.3 provides, in decreasing order of importance, the categories of image that the participants wished to keep private.

- **Personal and professional information:** Unsurprisingly, all participants wished to hide their private financial details such as credit/debit card details and cash or ATM card usage content if available in their lifelog data. The participants wanted to hide personal communication image data such as private emails, social media information, online website access or physical paper readings, personal messages over the phone, car plate numbers or personal bills e.g. electricity, telephone, shopping or grocery bills. Some participants were

Privacy assessment from number of participants (N = 25) for various types of content in visual lifelog data			
Content in lifelog data	No. of participants (Private (%))	No. of participants (Semi-Private (%))	No. of participants (Public (%))
Credit/debit card	25 (100%)	0	0
Private/professional emails on laptop/computer	25 (100%)	0	0
Cash money or ATM card usage	25 (100%)	0	0
Bathrooms	23 (92%)	2 (8%)	0
Personal communications on computer/mobile	22 (88%)	2 (8%)	1 (4%)
Images captured in a bedroom environment	20 (80%)	4 (16%)	1 (4%)
Self reflection in mirror	19 (76%)	4 (16%)	2 (8%)
Unflattering body positions	18 (72%)	5 (20%)	2 (8%)
Social media information	14 (56%)	10 (40%)	1 (4%)
Car speed/ driving style	11 (44%)	6 (24%)	8 (32%)
Paper/personal notes in physical medium	11 (44%)	10 (40%)	4 (16%)
Personal bills	10 (40%)	13 (52%)	2 (8%)
Car license plate number	7 (28%)	11 (44%)	7 (28%)
Personal physique, room or lifestyle	6 (24%)	15 (60%)	4 (16%)
Unknown people (Bystanders)	5 (20%)	4 (16%)	16 (64%)
Known person (Subject)	4 (16%)	18 (76%)	3 (12%)
Third person's house in images	4 (16%)	2 (8%)	19 (76%)
Clean home environment	3 (12%)	2 (8%)	20 (80%)
Pray	3 (12%)	2 (8%)	20 (80%)
Gardening at home	2 (8%)	2 (8%)	21 (84%)
General web surfing	5 (20%)	12 (48%)	8 (32%)
Any physical activity (exercise	4 (16%)	8 (32%)	13 (52%)
Taking rest in living room	4 (16%)	15 (60%)	6 (24%)
Food eating	4 (16%)	7 (28%)	14 (56%)
In outside environment with friends	2 (8%)	13 (52%)	10 (40%)
Normal driving images	1 (4%)	4 (16%)	20 (80%)
Shopping in grocery store	1 (4%)	3 (12%)	21 (84%)
Talking with family members	1 (4%)	20 (80%)	4 (16%)
Cooking	1 (4%)	12 (48%)	12 (48%)
Social interaction	1 (4%)	13 (52%)	11 (44%)

Table 3.3: Summary of the privacy assessment from the 25 participants of various types of content in visual lifelog data across three access levels: *Private*, *Semi-Private*, and *Public*, sorted in decreasing order of privacy concern.

working in professional roles and as researchers; these participants did not wish to reveal their professional information such as professional emails, meeting agendas, presentations and professional conversations with colleagues.

- **Environments (personal or private):** Most of the participants in this experiment wanted to restrict access to data that may reflect badly on themselves such as images taken in bathrooms, self-reflections in a mirror, unflattering body positions and messy/untidy environments. It is worth noting that this is quite a broad category and varies substantially across participants. Few participants not caring about messy environments but others caring a lot.

- **Identity of the person:** 16% of the participants wished to hide the identity of subjects i.e. people that they were directly interacting with, usually family members/friends/known person. 20% of the participants wanted to hide the identity of the bystanders as well i.e. people who are intentionally or unintentionally available in lifelog data without their consent.
- **Dietary routines:** Four participants did not wish to reveal their dietary routine with anyone. From their perspective, nobody would want to observe their food intake.
- **Other private information:** Some other types of private data that did not occur often in the lifelog data was still highlighted as private such as car speed/driving style e.g. hands on steering wheel, general web surfing, engaged in any physical activity or images taken in anybody's home besides their own.

2. **Semi-Private:** Table 3.3 shows the categories of images that participants were happy to share with family and friends in decreasing order of importance:

- **Family moments:** Most of the participants were ready to share images taken in the outside environment or inside their home with family and friends. For example, talking with family members (80% of the participants), having social interaction (52%) with friends or cooking in kitchen (48%) with family.
- **Personal lifestyle:** Some participants were happy to share the information about their personal physique, room or lifestyle with their family and friends. For example, talking with any known person (76%), revealing personal physique, room or lifestyle (60%), having a rest in the living room (60%), hanging out with friends or colleagues outside of the workplace (52%), exercising (32%) or doing some daily routine activities which would be well known by their family and friends (32%).

- **Using laptops:** Some participants (typically 48%) were happy to share images showing general web surfing, viewing media or TV, scrolling through social media content, watching movies etc. once the media does not contain personal or potentially embarrassing content.
- **Other information shared with family and friends:** Some other type of information available in lifelog images that was shared with family and friends such as personal bills (52%), car plate number (44%) and social media information (40%).

3. **Public:** This category includes any other visual lifelog image data that does not fall into *Only for me* and *Family/friends* categories above. This data is indicative of what the participants were willing to share publicly such as the visual content where user is gardening at home (84%), shopping in a grocery store (84%), images with a clean home environment (80%), content in images where the user is driving (80%) or engaging in worship/ praying at home or in church (80%), presence of an unknown person i.e. bystanders) (64%), etc. Some participants did not mind sharing images with food intake (56%), walking or doing exercise in a park (52%), cooking with family members (48%) etc. as discussed in detail in Table 3.3.

An example of lifelog images associated with each access level are shown in Figure 3.2. It should be noted that authors have applied these labels on the basis of the nature of our participant cohort, who were primarily academics or professionals. It is possible that other participants such as parents of young children may have different opinions or viewpoints. Authors expect that the general findings presented in this paper would remain similar.

By analysing Table 3.3, it is possible to draw conclusion about types of images that pose low, moderate or high privacy concerns for majority of participants. For example, authors observed that the images showing financial information or images taken in bathrooms carried a high degree of privacy concern for participants. At

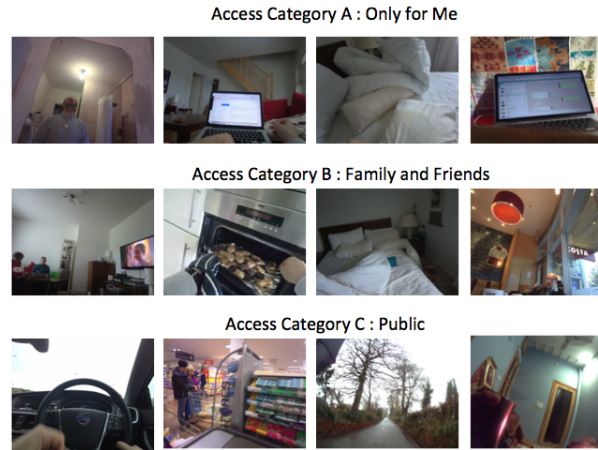


Figure 3.2: Example images from the specified three categories (i.e A, B and C) of anonymisation & highlight potential differences in willingness to share data i.e difference in ratio of bedroom images messiness from Access Category A: *Private* with Access Category B: *Family/ friends*

the other end of the scale, images showing driving, shopping or social interactions posted little or no privacy concern even if shared publicly.

Of course, this study also suggests an inherent difference between what individuals consider private. For example, some participants can share their dietary routine and their personal lifestyle publicly. But some participants wished to keep it private. Also, it is clear that there is a variance in tolerance of different levels of messiness in the home between participants. Even within participants, we can see a difference in tolerance. For example, two different images of the same bedroom in which one image with messy content is kept private while the other image with tidier bedroom is shared with family and friends as shown in Figure 3.2. Hence, it is important that individuals who gather such data are supported in easily defining what is to be shared and what can be kept private. In any case, it would be appropriate to err on the side of caution and enforce a strong privacy requirement.

3.3.6.3 Stage 3: Experiences of the wearers while lifelogging

Each participant was asked about their personal experience of lifelog data collection 10 days after completion of the experiment. It was the author's belief that this short separation between data gathering and interview session would facilitate a better understanding of the overall experience of the lifelogging. Findings suggest that more than half (56%) of the participants experienced an enhanced realisation about their daily lifestyle such as food habits, exercise routine or working hours i.e. self-realisation from two days of lifelogging. Additionally, almost half of the participants (48%) felt self-motivated and happy to review their lifestyle and activities, i.e. self-remembrance. A similar number of participants increased count of self-reminders to turn off the camera in bathrooms or while engaged in any intimate moments, i.e. self-control. Some participants (20%) felt that the lifelogging device affected their activities, i.e. influencing activities. For example, participants went for exercise in the gym regularly and started having healthy food in an organised manner. Six participants, all belonging to professional background were concerned to see so much confidential information such as ATMs or personal notes when reviewing their data i.e. self-agitation. Five of the participants who belonged to research community reported that data gathering was a normal part of their daily life and they did not even notice wearing camera. While one participant commented that he/she did not notice the device while busy in work but once outside of work there was an increased awareness of the device and an experience of difficulty in relaxing in social environments. However, most of the participants found that this experiment could provide insights for self-observance and self-awareness in daily routine, while some expressed surprise that they found it very useful for the self-observance in their daily life activities and behaviour.

3.3.7 List of Guidelines to Preserve the Anonymity

Based on the proposed experimental user study, we can recommend a list of guidelines for researchers who are working with real-world lifelogging data or releasing it publicly.

- We found a wide variance in privacy sensitivity between individuals. Hence, we suggested that participants gathering data for lifelog experiments should be given the opportunity to remove any sensitive data before sharing it with researchers, even if the data will be fully anonymised. This is in line with recommendations for using lifelog data in academic studies in [102].
- Due to the observations on the sensitivity differences among the cohort, we suggested that the participants in lifelogging studies should be involved in defining how/what data should be released to trusted parties. Should there be sufficient trust, then enhanced sharing may be possible, but the participant must be able to decide.
- De-identification of the data is a standard pre-release processing step for visual lifelog data [79]. It is not a panacea when sharing data and it will not solve the privacy problem for all lifeloggers. This is because semantic judgements on user context e.g. messy room are key factors and not the presence of faces or certain identifiable objects. Our understanding is that there is no suite of visual concept detectors currently available that are trained to identify some more semantically nuanced issues that we have found such as the difference between tidy and messy rooms, reflections, activities in the bathroom, etc.
- Trusted-researcher agreements that clearly outline the expectations on researchers to respect the privacy of donating lifeloggers and implement appropriate data governance methodologies are also highly recommended. This is the process that was followed for the NTCIR - 12, 13, 14 lifelog datasets that were released in recent years [108].

By following these guidelines (explained in [103]), authors motivate research community to understand privacy concerns of lifeloggers who gather lifelog data to share with family and friends and with the third party. In line with the previous work [104, 109], authors investigated and analysed the sensitive content available in the visual images that stated the sharing preferences of participants (results shown in Table 3.3 earlier). Unlike the previous studies [104, 109], where visual lifelog data was captured from a small group of university students over a short period of time, authors have captured 50 days of visual lifelog data (in total) from 25 participants with a diverse profession. Authors found a significant difference between participants in terms of what they are willing to share or keep private as summarised in Table ???. Authors found some results different from Hoyle et al. [109] and Chowdhury et al. [104]. In particular, 48% participants who were mostly university students and early stage researchers in this study were happy to share images showing general web surfing, viewing media or TV, scrolling through social media content, watching movies etc. if media does not contain personal or potentially embarrassing content. This would differ from the more experienced professionals who were comparatively more privacy-aware for sharing their lifelog data. Although it was a preliminary study, it allowed the community to draw some initial conclusions which helped us to generate privacy-aware lifelog datasets for release or for lifelog data segmentation user studies. We refer [103] and appendix A.4 to the readers for better understanding of the findings of this study and major contribution of the dissertation writer in this user experiment. We now present the process of releasing privacy-aware lifelog dataset that automates the anonymisation process as described above.

3.4 Releasing Privacy-Aware Lifelog Test Collections

In past lifelog researches, experiments were often done on small datasets that were not released for comparative use. In few cases, the researchers themselves gathered the datasets. It is not ideal as discussed earlier in Chapter 2. Developing a large

volume of lifelog data from willing lifeloggers and making it available to the public is very challenging. It is our belief that gathering comparative and reusable nature of lifelog datasets benefits research community. Therefore, we decided to create and use the reusable publicly available privacy-aware lifelog datasets for this research.

The datasets we have gathered to support this research requires a number of processes to get it completed such as structured data gathering by a number of individuals over an extended period of time, removal of user-identifiable content and organisation of data hierarchically into basic minute-long intervals, dataset publishing with setting up individual and organisational data agreements, definition of the ground-truths and evaluation criteria and definition of the evaluation topics i.e. user queries. By understanding the process of generating lifelog test collections as explained in [78], we suggest a number of principles for organising and releasing lifelog data to the public.

3.4.1 Principles of Generating Lifelog Dataset

Following NTCIR-12 [10, 99] and the LTA-2018 (introduced in Chapter 2), we propose required principles which could create uniformity in evaluation tasks to target similar tasks over years with comparative and repeatable test collections. These principles resolve the major challenge of how willing are users to share data. We pointed out previously while we ourselves and some other volunteers provide the solution for the reported challenge. Thus, we discuss all the suggested principles as follows:

- **The Continuity:** Lifelog data of each individual should be captured continuously over a number of days in terms of recollecting and recalling intentions supported by Sellen et al. in [17]. Getting meaningful information from lifelogs could be a retrieval problem e.g. “find the moments that a logger having dinner” or could also be an analytics problem e.g. “compare the eating habits between two loggers”. In order to answer these questions, the dataset must contain sufficient information i.e. the lifelogs of each logger

should be gathered continuously over a number of days. In our initial user experiment which will be discussed in Chapter 5, we collected lifelog data from 10 users over the period of one day each. Later in this study, we used a publicly available lifelog search challenge (LSC2018) test collection to validate our preliminary results to the research community which consisted sufficiently large volume of lifelog data. It typically comprises 27 days of lifelog data from one active lifelogger. We observed that the data captured over a longer period of time is sufficient for identifying habit patterns of an individual in this study.

- **The Completeness:** In this research, we intend to digitally capture the totality of the life experiences of an individual. From past research work in lifelogging as explained in Chapter 2 [1, 41, 42, 50, 51]) typically used visual data and biometric data. Gurrin et al. in [4] presented different sources of data gathered in lifelogging domain. They introduced a collection of primary data that includes source such as physiological data from wearable sensors e.g. heart rate, respiration rate, galvanic skin response, etc., movement data from wearable accelerometers, location data, nearby bluetooth devices, WiFi networks and signal strengths, temperature sensors, communication activities, data activities, environmental context, images or video from wearable cameras. It does not take into account in terms of semantic metadata. All these are vastly varied data sources. In lifelogging, all these varied sources merge and combine together to form a holistic personal lifelog where variety of data sources is normalised and eliminated.

In the case of lifelogging, because much of our lifelog data can be derived from sensors which may be troublesome or sometimes being imprecise and uncertain or have calibration and drift issues as described in [110]. We investigated that the lifelogs should contain four basic types of information: visual data, personal biometrics, human activity, and information accesses as

explained earlier in this chapter. In order to allow statistically significant study in lifelogging field, a dataset needs to be large enough to have sufficient amount of required sources of data for the target analysis to represent real-world data of lifeloggers [99].

As addressed in the study above [4], we motivated to log primary information to gather lifelog data. We propose these four sources [78] as primary information data that a lifelog dataset should contain: **(i) Visual information:** use for continuous and automatic capture of life activities as a sequence of digital images. **(ii) Personal biometrics:** the lifelogs should contain basic personal biometrics of the individual e.g. heart-rate, calories burned, etc. **(iii) Human activity:** there are many personal sensing devices such as accelerometer and location data for monitoring everyday activities which can monitor human performance. For example activity levels such as number of steps, distance travelled, sleep duration, etc. for the recall and retrieval purposes [17]. **(iv) Information access:** this could refer to using a computing device e.g. smartphone or computer to actively and passively capture the user's context or logging the consumed data from all words typed, web pages read, videos watched and so on. We identified the need of capturing human computer interaction data such as emails, websites visited, mouse and keyboard tracking information, periodic screenshots, etc. Authors believe that these four categories of data collection are sufficient for analysis of an individual's life experiences.

- **The Anonymity:** All personal identifiable information (PII) data must be removed. We must adhere to the principle of privacy when creating the lifelog dataset from our user study explained in previous section [103]. we propose data anonymization to remove PII data while maintaining the usefulness of the data. Human faces, personal IDs, vehicle plates, name tags, and address location of the individual's home and work place explained earlier in this

chapter are considered as examples of PII data.

- **The Protectiveness:** The dataset should be password protected and all accesses should be logged. The lifeloggers who donated the data or any identifiable individuals who appear in the data should be able to request some content to be deleted at any time. The data should be protected strictly by agreements between loggers, users, and the people who manage the dataset.

3.4.2 Process of Releasing Lifelog Test Collections

In order to build a lifelog dataset that respects the proposed principles as described earlier in this section, we apply a 3-step process as follows. Initially we gather the lifelog data from individuals by following the first two principles i.e. continuity and completeness. Next, we clean the gathered data by removing PII data by applying the anonymity principle as explained earlier. Finally, before publishing the dataset, we protect the data by enforcing agreement between loggers, users, entities who manage the dataset. It is further enhanced by putting in place a data download tracking mechanism. These steps are described as follows:

- **Lifelog data gathering:** Due to the long-term, always-on nature of lifelog data gathering, it is important to reduce the overhead on the lifelogger of gathering the data. Hence, the data should be gathered using wearable devices, for example autographer, narrative clip, and/or a smartwatch (see Figure 3.3). It is worth noticing that in LSC2018 dataset, an active lifelogger gathered continuously over 27 days which fulfilled our first principle of continuity in capturing lifelog data.
- **Lifelog data preprocessing:** From data gathering, second step is to extract the visual features from visual lifelog data to address the second principle of completeness. After extracting features from lifelog data, we remove the personal identifiable information (PII) data by applying anonymisation process

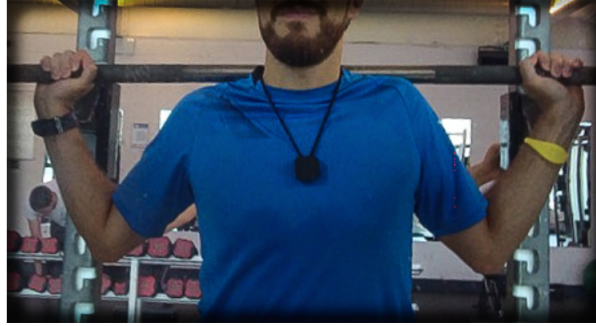


Figure 3.3: A lifelogger with his everyday wearable devices: A Narrative camera (wearing on his neck) and a smartwatch (the black watch on his right arm).

and techniques as discussed in the previous section which addresses the third and fourth principle of anonymity and protectiveness.

- **Lifelog data publishing:** For publishing dataset, we set up agreements between lifeloggers, users, and entities e.g. researchers and organisations who manage the dataset and keep track of all accesses as explained earlier in this chapter while discussing users and their issues. We also define two stakeholders and two agreements of the published dataset as follows:

- **Lifeloggers (stakeholder):** provide the data and also do the self-cleaning on the data. The lifeloggers also take the responsibility of letting the individuals in their field of view be fully aware that the camera is capturing the individuals.
- **Researchers (stakeholder):** who manage the dataset: do the anonymisation and organise the data.
- **Data storage, privacy access, and agreements (agreement):** Beside hosting the dataset, this “component” also protects the data by applying password protection of dataset. We also protect the data by making individual and organisational data agreements.
- **Data deletion policy (agreement):** the lifeloggers and individuals in the dataset can request data deletion any time and need to agree to use of data beyond the initial plan.

Thus by following this process, we have released privacy-aware lifelog test collection to offer researchers that can be used for various information retrieval tasks in lifelogging. This step resolves our second research question of construction of privacy-aware lifelog test collections to support comparative benchmarking for ad hoc retrieval tasks addressed in Chapter 1. We will describe the lifelog datasets used in this research in the following section.

3.4.3 Lifelog Test Collections Used in this Research

For this research, we mainly used the publicly available lifelog search challenge (LSC2018) test collection [78] while evaluating our proposed segmentation algorithms as will be explained in Chapter 5 and 6. LSC2018 is a subset of NTCIR-13 lifelog dataset, which consists 27 days of multimodal lifelog data, typically 38,880 minutes/moments, generated by one active lifelogger. Various forms of metadata, visual features (tags) using Microsoft cognitive services API [3], manual annotations, and various other data sources that capture the real-world activities of the user as also present and associated with this visual data. We have released additional manual annotations as an addendum to the LSC2018 collection in July 2019⁴. Additionally, we have used NTCIR-12 and NTCIR-13 lifelog datasets [8, 10] for our supportive experiments, [8, 111].

Table 3.4: The summary details of all the publicly available datasets along with other metadata being used for this research.

Type of Content	LSC2018	NTCIR13	NTCIR12
visual images	y	y	y
song history	y	y	n
light status	y	n	n
biometric data	y	y	n
weather status	y	n	n
date and time	y	y	y
manual food log	y	y	n
manual drink log	y	y	n
semantic locations	y	y	y
physical activities	y	y	y

For our preliminary user experiments as outlined in Chapter 5, we collected

⁴LSC2018 Dataset available at: <http://lsc.dcu.ie>.

short-term lifelog data from a number of individuals to support initial experiment of building privacy aware mechanism for lifeloggers. We found the best approach to lifelog segmentation using this data. In these experiments, the wearable camera data is collected from 10 people for the duration of 1 or 2 days each e.g. 12-14 hours in a day and this data was enriched by extracting visual features using Caffe framework [2] and concepts from the Microsoft cognitive services API [3] as introduced in Chapter 2. The data is gathered by using autographer wearable lifelog device, which can automatically capture up to 2500 images per day. Table 3.4 summarises all the datasets along with other metadata we used for this study. It is worth noticing that following the previously described principles, LSC2018 and NTCIR-12, 13 - Lifelog datasets can support a number of years of ongoing research activities.

After collecting privacy-aware lifelog data from selected participants, one can move to the next step of the proposed test collection methodology i.e. organising the collected data in the structured format for analysis. We need to synchronise and align the multi-sensor lifelog data into some sort of indexable unit that would provide meaningful insights to the user. In the next section, we introduce the units of retrieval in lifelogging and will provide the rationale behind choosing the most appropriate unit of retrieval for this dissertation.

3.5 Basic Units of Retrieval

To provide useful and indexable content for a lifelog retrieval system, it is necessary to align and synchronise streams of multimodal lifelog data and combine them into some form of coherent unit. This unit is typically called a document. However, there has been little consideration of what a document means for lifelog-based retrieval [11]. Therefore, we define a new hierarchical model for unit of retrieval in lifelogging (shown in Figure 3.4) as follows:

- **Item:** An Item is defined as the atomic retrievable unit such as an image,

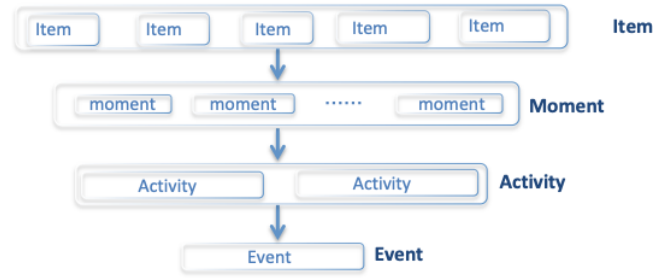


Figure 3.4: Hierarchical Model to define basic units of retrieval in lifelog data based on the intentions of recalling and recollecting personal life experience.

temperature reading, location, etc. (see Figure 3.4). It is the unit of retrieval that was favoured in MyLifeBits [29] to generate the next generation storage systems that make the captured material such as video, audio, images, web pages, email text, other text documents etc. manageable and useful rather development of a software system.

- Moment:** Moment is defined as a fixed length temporal unit which heretofore has been a minute. There are 1,440 moments in each day and the moment is represented by a combination of all the atomic items that take place within that moment (see Figure 3.4). Moments were used as the retrieval unit in the lifelog semantic access (sub) task at NTCIR- 12, 13, 14 lifelog comparative benchmarking workshop (which we will detail later) [8] in which the focus is to retrieve a number of specific moments based on a specific query as per user information need in a user's life that happened throughout the day.
- Activity:** Activity is defined as an un-interrupted sequential state of human activity in terms of their individual environment or stimuli. Activity is the indexing-time unit of retrieval that we define in this research work and propose as the most appropriate indexing unit. It represents a combination of sequential moments (see Figure 3.4) whose size depends on the duration of activities of the individual and dynamic.
- Event:** Event is a combination of moments and/or activities segmented at

indexing time. Typically, an event is more fixed and static in nature that represents the longest unit. There is typically 2-4 events in any given hour or typically 1 event over the duration of 3-4 hours or more based on past research [1, 12]. For instance, the individual is watching a movie for 3 hours or engaged in any family function e.g. birthday party or marriage ceremony for longer than 5 hours.

Since lifelogging is an inherently multimodal data capture activity, such an indexable unit would manifest as a fusion of a number of synchronised 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. This unit would represent a contiguous period in time and would form the basis of a document of life experience data. Providing access to a large number of such documents would form the basis of a lifelog retrieval engine.

Besides test collections, other researchers have considered the reasons for accessing lifelogs. We propose that it is necessary to take a step back and explore many potential ways that an individual in daily life can use lifelogs. Sellen and Whittaker [17] suggested five R's of memory access as reasons why individuals access their memories. By extrapolating from these five Rs, it can be identified five reasons why an individual might keep a lifelog. The 5 Rs 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.

We are focusing on recalling and recollecting as the most appropriate forms of retrieval from lifelogs to build a system that would support a user to store, encode and access his/her personal lifelog data with ease in this research. It is our conjecture that neither moment nor event units are suitable to provide a flexible retrieval system for the user which can address these two Rs- recalling and recollecting;

defined as focus of this dissertation in Chapter 1. Defining the minute as the unit is clearly too small to represent a human activity in terms of where user needs to retrieve the time when individual is working on laptop or watching TV at home. Activities performed consciously or unconsciously for a minute such as door slam or picking up the pencil from floor do not provide any meaningful information on their own. Whereas defining the unit as an event could result in units of retrieval that are of fixed duration and too long to adequately address an information need. For example, consider an event such as a birthday party, there could be many user activities overlapped in one large event. We identified that the minutes are too short and events are too long and static in nature to retrieve meaningful information. Therefore, a new unit called an activity is introduced in this dissertation which proposes to be a more suitable unit of retrieval i.e. see Figure 3.4. Activity as a unit is likely to be more dynamic in duration than an event and more robust to satisfy user information need in terms of user activity than a minute.

3.6 Evaluation Measures and Techniques

Based on our literature review (in Chapter 2) [1, 5, 12, 16], we have observed that most segmentation algorithms used the sliding window approach. It is performed on required operation on specific window size of given time-series data where window starts from the first moment and keeps shifting right by one to evaluate time-series data. In lifelogging, the captured multi-sensor lifelog data is continuous in nature and we need a flexible mechanism to evaluate the accuracy of the detected activity boundaries by proposed segmentation algorithm.

3.6.1 Evaluation Metrics

After implementation of sliding window approach, we get the similarity matrix for the detected activity boundaries against user ground-truth activity boundaries. In terms of evaluation, we then turn to the conventional measurements of precision,

recall, F1-score, and Matthew's Correlation Coefficient (MCC). Precision and recall are the standard approaches for the evaluation measurement in information recall. F1-score is a harmonic mean of precision and recall. Matthew's Correlation Coefficient (MCC) score considers not only the correct prediction but also measures the correlation between all values of the score vector. It identifies the missed predictions by adopting values smaller than 0 which presents the classification results in more detail. All these four measures are used to present the overall performance of proposed approaches to lifelog segmentation algorithms.

		Ground Truth		
		Positive	Negative	Total
Predicted	Positive	TP	FP	TP + FP
	Negative	FN	TN	FN+TN
Total		TP + FN + FP + TN		

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{Precision} = \frac{TP}{TP+FP}$$

$$\text{F1- Score} = \frac{2.TP}{(2TP+FP+FN)} \quad \text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

$$\text{MCC} = \frac{TP.TN - FP.FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

Figure 3.5: Confusion matrix: A standard metric for segmentation and classification problems.

We also used confusion matrix as a standard approach to evaluate system performance as shown in Figure 3.5) to get the deeper analysis. True positives i.e. where both user and system defined segmentation points are exactly matched. True negatives i.e. where no user and system-defined segmentation points are detected. False positives i.e. where no user-defined segmentation point detected, but the system defined segmentation is detected. False negatives i.e. where a user-defined segmentation point is detected, but no system defined segmentation point is detected.

3.7 Summary of Proposed Approaches

To make multi-sensor lifelog data understandable to the users and to address all the research questions described in Chapter 1, we have employed the methodology proposed (in section 3.1). We proposed and implemented different approaches to resolve each research question (shown in Table 3.5). As we can observe from our initial steps in proposed methodology, one needs to build privacy-aware reusable lifelog test collection. Therefore, we have presented the user studies to resolve these issues which also answers first and second research questions. Lifelog data usually gathered in raw unstructured format from various different sensors and it is our belief that we need to synchronise and align such multi-sensor lifelog data. We discuss the process of generating lifelog documents in Chapter 5 in detail and it also addresses our third research question. Later in this study, we performed one initial user experiment with 10 users who collected lifelog data over the period of 1 day each and defined our baseline segmentation approach for visual lifelog data. To measure the quality of the state-of-the-art approach for this dissertation, we implemented and evaluated conventional static segmentation algorithm using publicly available lifelog search challenge (LSC2018) test collection (will be discussed in Chapter 5). We took a step further and tried different segmentation algorithms by using LSC2018 lifelog test collection to improve segmentation at query time. Also, we have proposed a methodology to segment lifelog data both at indexing and retrieval time. This will be discussed in detail in Chapter 5 and 6. Finally, the outcome of proposed methods proved to be promising and motivated us to compare segmentation algorithms for both static indexing-time and dynamic query-time approaches to show how dynamic approach outperforms the conventional static segmentation approaches. This will be discussed in Chapter 6. The experimental configuration for all tasks in this study is summarised in Table (3.5) below:

Table 3.5: Summary of experimental configuration for this research.

Chap	RQ	Experiment	Dataset	No. of Users/Days	Main Contribution of the Experimental Study
4	RQ 1	Data Anonymization	self gathered	25/50	Variance of privacy concerns across participants
4	RQ 2	Test Collections	NTCIR-13	2/45	Process of releasing privacy-aware lifelog dataset
5	RQ 3	Document Generation	LSC2018	1/27	Process of generating documents in lifelogs
5	RQ 4	Static Segmentation	self gathered	10/10	Basic segmentation approach using visual lifelog data
5	RQ 4	Static Segmentation	LSC2018	1/27	State of the art segmentation approach using multimodal lifelog data
6	RQ 5	Dynamic Segmentation (Moment)	LSC2018	1/27	Using mAP score to evaluate LSAT topics
6	RQ 5	Dynamic Segmentation (Activity)	LSC2018	1/27	Comparing Static and dynamic segmentation using LEST topics
6	RQ 5	Dynamic Segmentation (Manual Ann)	LSC2018	1/27	Improving performance of dynamic segmentation using 24 new LSC topics
7	RQ 5	Comparison Static v/s Dynamic	LSC2018	1/27	Identify better segmentation algorithm

3.8 Chapter Summary

In this chapter, we described the methodology followed for this research work. To conclude this chapter, we presented our experimental configuration to build information retrieval systems. We presented the primary motivation to build lifelog test collections, discussed various challenges faced while collecting lifelog data, and listed the main data sources we have used in this dissertation. The readers can see the main contributions of the dissertation writer for this study in appendix A.4. Lifelog studies always involve a rich amount of multi-sensor personal archives which raises serious privacy concerns [112]. Consequently, we investigated the process of selecting lifelog users and their associated privacy issues. We also provided the solutions to resolve participant’s privacy issues in this chapter. We presented a small user experiment to show the need of anonymisation of lifelog data before releasing publicly. We proposed a hierarchical model to identify the basic units of retrieval in the lifelogging where we presented the rationale about how we decided the most appropriate basic unit of retrieval for this dissertation. In addition, we presented the overview of evaluation methods and techniques used for this research work. Lastly, we summarised all the proposed approaches to resolve all the research questions described in Chapter 1 that will be discussed in detail in forthcoming chapters.

Chapter 4

Lifelog Data Processing and Data Analytics Techniques

4.1 Introduction

In a previous chapter, we presented a methodology for building lifelog test collections. We also proposed the process of releasing privacy-aware lifelog test collections for comparative evaluation. In this chapter, we follow the proposed methodology and focus on discussing lifelog data pre-processing techniques which are used in current research. Image data is the primary source of data used in this research work as discussed in previous chapters. Process of cleaning lifelog images is introduced in section 4.2. Various data analytic techniques such as vector distance methods and thresholding have been used to analyse lifelog data in this work are also presented in section 4.3. Lastly, in section 4.4, we conclude this chapter by discussing all the techniques we used to process lifelog data in this dissertation. Lifelog data processing techniques used in current work are explained in the following section.

4.2 Lifelog Data Preprocessing Techniques

In this section, the process of lifelog data pre-processing is presented which is an important step to train, validate and test any proposed segmentation algorithm. We follow steps of pre-processing lifelog images i.e. checking the format (grey scale

or RGB) of the images, resizing, removing any noise and extracting visual features from lifelog images. Lifelog images are collected using various wearable devices and are inconsistent with basic properties of images such as image, shape, size and consistency of noise level. Therefore, it is important to clean lifelog image before analysis. Steps required for lifelog data pre-processing is explained in following sub-sections and the implemented algorithm is shown in Figure 4.1.

4.2.1 Checking the Shape of Lifelog Images

An image is stored as a matrix of pixels and each pixel contain different number of channels. For example, a grey-scale image has only one channel i.e. it has shades of grey colour between 0 to 255 where 0 corresponds to black and 255 corresponds to white. A coloured image has three channels i.e. red, green and blue where each pixel has the value of RGB output signal between 0 to 255. Therefore, for ensuring consistent data cleaning process, all the images should be converted to a standard format.

4.2.2 Resizing Lifelog Images

In lifelogging, images are being captured automatically or manually using various wearable cameras and phone cameras. It should be noted that the size of images could vary considerably depending upon the device. For lifelog user interfaces such as query search engines, we define 1024 X 768 the standard size of an image. However, we needed to reconstruct images to establish a base size for maintaining uniformity for all the lifelog images in the corpus for segmentation task. In this work, we usually resized all images to (220, 220, 3) dimension for visual analytics. The first parameter represents height with the value of 220, second parameter represents width with value of 220, and third parameter represents RGB channels with value of 3 (see Figure 4.1).

4.2.3 Removing Noise from Lifelog Images

De-noising the lifelog images is one of the major pre-processing step. It should be noted that the digital images must have some noise that should be reduced using noise reduction filters such as Gaussian blurring or Median blurring. This step helps in better analysis of the proposed segmentation algorithms (see Figure 4.1).

```
#Checking the shape of images
1. Define a function
2. INPUT: dataset
3. READ: images in dataset
4. FOR n in dataset
    IF n ends with .png THEN
        RETURN image shape with height, width, and channels
    ENDIF
ENDFOR

#Resize images
5. Define the size of image with height (220), width (220), and channel (3)
6. FOR i in dataset
    RESIZE every image in dataset
ENDFOR

#Denoising images
7. Define a empty array
8. FOR i in resized image dataset
    Compute GaussianBlur method to remove the noise
ENDFOR
```

Figure 4.1: Pseudo-code for lifelog data pre-processing.

In image processing, a Gaussian blur also known as Gaussian smoothing is blurring an image by a Gaussian function to enhance image structures at different scales. Median blur is a non-linear digital filtering technique which is typically used to preserve edges of objects in the images while removing noise for next step of data processing, e.g. edge detection on an image.

4.2.4 Feature Extraction and Concept Detection

Over the last decades, one of the most popular multimedia retrieval workshops such as ImageCLEF¹ (introduced in Chapter 2) have used multimodal lifelog data and supported visual concept detection and object annotation tasks in visual analytics

¹<https://www.imageclef.org/2008/vcdt>

for the research community. From previous studies [1, 12, 36, 78, 80] (explained in Chapter 2), the researchers typically used pixel level attributes such as colour layout, colour description, edge histograms, etc. We found the need of semantic-level visual concepts for extracting semantic labels for each object in visual data. This helps to improve the accuracy of the retrieved information. We have used Caffe framework, [2] as introduced in Chapter 2 to extract visual features from lifelog images. Caffe visual concepts classify an image into 1000 categories of ImageNet classification. We also explored other visual concept detection tools which can provide more detailed semantic labels for image content. We also extracted visual features using vision recognition and content detection API introduced by Microsoft cognitive services [3]. The examples of both concept detection tools will be provided later in this chapter. In this section, we explain automatic visual concept detection tools [2,3]; supportive metadata such as activity based manual annotations and contextual and descriptive data used in this research.

4.2.4.1 Automatic Concept Detection Tools

- **Caffe visual concepts:** Caffe stands for ‘Convolutional Architecture for Fast Feature Embedding’. It is based on deep learning framework [2]. It is based on Convolutional Neural Network (CNN). It is open source and provides python command line or MATLAB user interface. It supports many types of deep learning algorithms geared towards image classification and image segmentation. We used python interface of this deep learning classifier and trained our collected lifelog data. This classifier outputs label and its probability of occurrence for 1,000 objects based on ImageNet classification for an image. For an instance, we can observe visual features and their related confidence scores for each object in every image (see Figure 4.2).
- **Microsoft cognitive services API:** Microsoft cognitive services [3] provide a number of APIs such as vision APIs to detect semantic information from image e.g. facial recognition and recognise digital ink in the content;

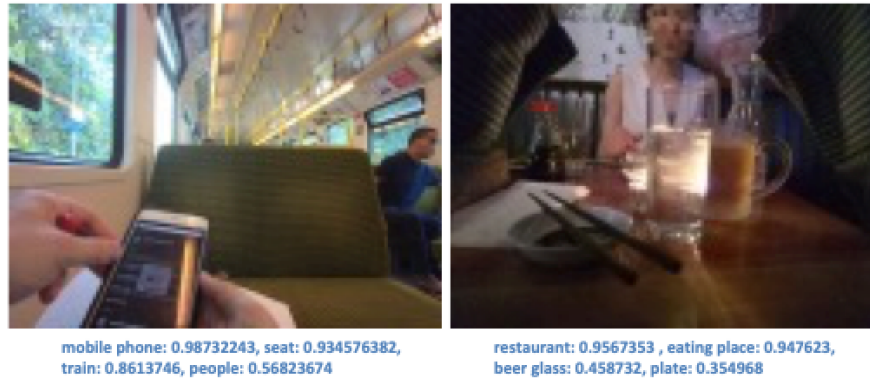


Figure 4.2: Example of lifelog images with extracted visual features associated with confidence values.

Speech APIs for speech services such as speech to text, text to speech, speech translation and speaker recognition; and language APIs for text analysis. In this study, our primary focus is visual lifelog data. So we used vision recognition and content detection by API² for this research work. As discussed in Chapter 2, computer vision API provides different visual features such as image caption, categories, tags, adult racy score, description of image. However, we observed image categories and image tags are well explained to classify each object, building or person which can benefit from the segmentation task. Therefore, we extracted and used visual concepts such as tags i.e. tagging each object in a particular image associated with confidence scores i.e. which are based on thousands of recognisable objects, living beings, scenery, and actions; and categories based on parent-child hierarchy where it provides 86 super categories and each super category is associated with sub categories of various objects. For example, animal category has further sub categories such as bird, lion, bear, etc. Examples of other categories are abstract: abstract_shape, animal: animal_bird, building: building_church, dark: dark_fireworks, drink: drink_can, food: food_pizza, indoor: indoor_churchwindow, outdoor: outdoor_waterside, object:

²<https://docs.microsoft.com/en-us/azure/cognitive-services/computer-vision/home>

object_screen, others, people: people_baby, plant: plant_flower, sky: sky_cloud, text: text_map and trans: trans_car.). For an instance, we can observe the semantic labels for a lifelog image using computer vision API (shown in Figure 4.3).



Figure 4.3: Example of visual features extracted using vision recognition and content detection API [3].

It should be noted that although it reduces the human effort of understanding image content, it may not solve the semantic gap in the content-based image retrieval i.e. pixel-level image reading to semantic-level image labels. To bridge this gap, we used manual annotations that will be explained in next subsection.

4.2.4.2 Manual Annotations

In lifelogging, manual annotations include human annotations of visual lifelog data in terms of semantic labels that could help in retrieving meaningful information. Lifelog data contains very sensitive and private content. We still believe that manual annotations should be undertaken by expert trusted researchers who have enough understanding of lifelog data to preserve privacy of the lifelog data and to maintain the uniformity of lifelog data for the target analysis. Such a researcher should know the target use case of developing lifelog retrieval systems; the researcher should have a proper understanding of lifelog data; and researcher needs to decide whether the annotations should be more precise or in detail such as

eating food and eating pizza, apple or burger specifically. Although this is more labour intensive task than automatic annotations discussed earlier, it facilitates naturally more accurate annotations than the automatic annotations in terms of detailed human understandable labels. Therefore, we used manual annotations in our research. We build single level activity-based ontology for manual annotations to improve accuracy of the proposed segmentation algorithms as discussed in detail in Chapter 6.

4.2.4.3 Contextual and Descriptive Metadata

Metadata is described as the data that provides the information about the actual data. Content of metadata could be descriptive, structural, contextual or statistical information of the actual data. We used contextual and descriptive metadata to describe an image content. This could be beneficial for search and discovery contexts. This metadata drives the ability to search, browse, sort and filter information in ad hoc search engines. We extracted metadata from the captured lifelog data images such as date and time from image and sensor metadata; personal manual context including age, gender, height, weight, mood, food, drink logs using mobile phone applications, semantic locations including home, restaurant, work using moves app; environmental context including weather status, day status. Extracting automatic and manual annotations along with contextual and descriptive metadata generates a lifelog test collection for the comparative evaluation in information retrieval tasks. The example image of using automatic annotations (using visual concept detectors), manual annotations and contextual and descriptive metadata is shown in Figure 4.4.

We followed the steps mentioned above for lifelog data pre-processing. One can then implement machine learning methods or data analytic techniques to retrieve meaningful insights from lifelog data. We explain the lifelog data analytics techniques we have used in the following section.

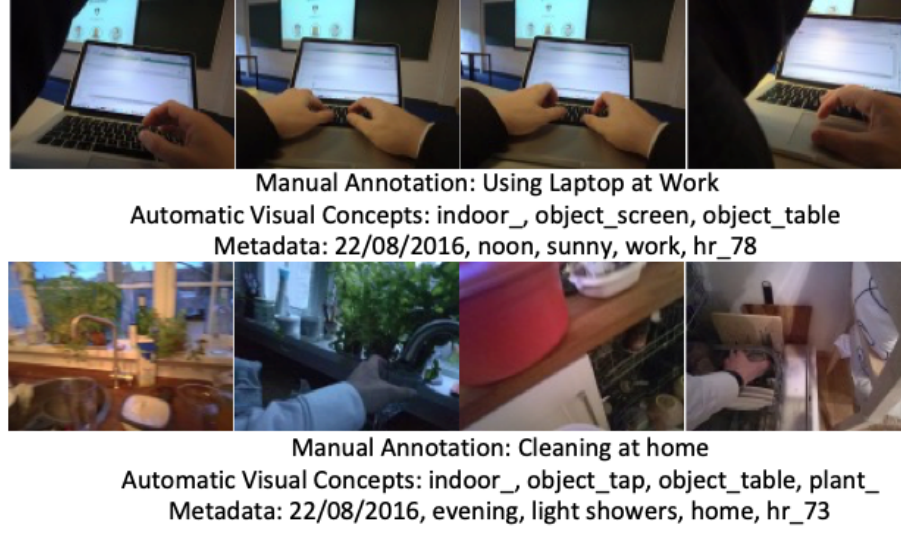


Figure 4.4: Example image with extracted automatic visual features, manual annotations, and metadata.

4.3 Lifelog Data Analytics Techniques

We have used various data analytic techniques to extract semantic meaning of the multimodal lifelog data. For instance, the GPS on the phone captures the location data in terms of longitude and latitude values. Data analytics techniques help to convert these numerical values into semantic names of the locations such as work, home, restaurant, church, etc. Similarly, we convert time data into day light status data. Time is one of the prime source of retrieval for search and retrieval purposes, but inefficient to search information using exact time or specific time range. Therefore, we use day light status such as early morning, afternoon, late night which is inherently generic to search and retrieve information. Besides datasets and evaluation criteria as described in Chapter 3, there are several data analytic techniques that we used for this research as discussed in following subsections.

4.3.1 Vector Distance Methods

While implementing machine learning algorithms, vector distance methods were used to observe the input data patterns to make any decision. Vector distance

methods help to improve the performance of classification, clustering, segmentation and information retrieval process significantly. Vector distance function is based on a Pythagorean theorem that considers distance between two data points. In lifelogging, various vector distance measures such as Euclidean vector distance method, Manhattan distance, cosine distance or histogram intersection distance method could be employed to segment a continuous stream of lifelog data to recognise the user activities performed in a day. To achieve this, we implemented Euclidean vector distance method. It has proved to be the best vector distance method based on our baseline segmentation approach [1] that will be discussed in Chapter 5. It calculates distance between every pair of adjacent images based on confidence scores of the visual features in time-series photo stream data, see Formula 4.1).

$$Euclidean_Distance(V) = \sqrt{\sum_{n=1}^m (v_2 - v_1)^2} \quad (4.1)$$

By observing the Euclidean space, we can identify change in activities having high distance in two adjacent images e.g. image transition from having breakfast at home to driving a car) and fuse the adjacent images with small distance e.g. talking over phone while driving a car. It detects change in activity boundaries based on Euclidean distance between images.

4.3.2 Thresholding Techniques

We used both automatic and manual thresholding techniques to highlight top ranked activity boundaries. We implemented two thresholding methods. First is automatic mean thresholding which is a parametric technique based on mean, standard deviation and user parameter K, where K could be initialised at standard value 0.5 at the initial state and then could be changed to any other value to optimise the overall system accuracy. Second is Kapur thresholding which is a non-parametric technique based on entropy [1]. Refer Figure 4.5 for implementing

Euclidean vector distance algorithm and automatic mean thresholding method. Further discussion about thresholding techniques will be done in following chapters.

```
#Implementing Euclidean Distance
1. Define a function
2. INPUT: vector of confidence scores for each image
3. Define a empty vector
4. FOR i in vector
    Compute the Euclidean distance
    RETURN distance
ENDFOR

#Implementing Mean Thresholding
5. Define a function having distance vector
6. Calculate the mean
7. Calculate the standard deviation
8. Initialise the parameter K
9. Calculate the mean threshold using mean_threshold = mean * K + std_deviation formula
10. Check where distances are greater that mean threshold value
```

Figure 4.5: Pseudo-code for calculating the Euclidean vector distance method and mean thresholding method.

4.4 Chapter Summary

In this chapter, we presented the techniques for lifelog data enrichment and pre-processing. This chapter presented various techniques to extract the visual features from visual lifelog data, introduced manual annotations and contextual/descriptive metadata. In addition, we introduced the lifelog data analytic techniques used in this research work. In the next chapter, we will define the process of generating lifelog documents, which address our third research question (RQ3). We will identify the most appropriate basic unit of retrieval for lifelogging data for this dissertation. We will then explore the existing state-of-the-art approaches to define our baseline segmentation approach and will identify the best state-of-the-art static segmentation approach at indexing-time in the next chapter. This will address our fourth research question (RQ4) as described in Chapter 1.

Chapter 5

Static Segmentation of Lifelog Data

5.1 Introduction

In previous chapter, we explained methods to pre-process and analyse lifelog data. In this chapter, we explore existing state-of-the-art approaches to temporal event segmentation and our own advancements in lifelogging. This chapter focuses on generating lifelog documents where a document is logically related temporal subset of consecutive minutes comprising multi-sensor data. This document is a retrievable unit as introduced in Chapter 1. This chapter explains baseline segmentation approach [1] as reported in Chapter 2 and proposes two new approaches of temporal event segmentation of lifelog data at indexing time. We identify the best approach to temporal event segmentation. It will be the basis for the comparison with segmentation approach based on activity retrieval unit at a later stage in this chapter. We explore the process of generating lifelog document which addresses our third research question (RQ3) i.e. what are the various units of retrieval documents for ad hoc search through multimodal lifelog data?.

5.2 Events are Documents

There has been a minor consideration of what a document means for lifelog-based retrieval. To provide useful and indexable content for a lifelog retrieval system we

have used a six-step process for generating lifelog documents as shown in Figure 5.1.

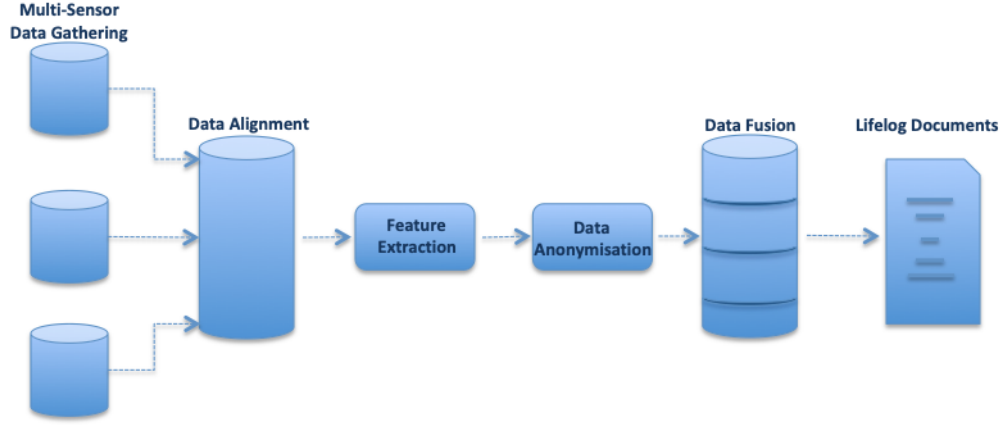


Figure 5.1: Process of generating lifelog documents in lifelogging.

- **Data Gathering:** Initially, we gather multimodal lifelog data from several sources such as audio, video, images, biometrics and locations as discussed in [4]. We have presented the process of gathering lifelog data in Chapter 3.
- **Data Alignment:** This multimodal data has different levels of granularity in terms of volume, fidelity, accuracy and semantic meaning. Ultimately, these vary from one lifelogger to another as discussed in Chapter 3. These continuous multimodal sources of data must be temporally time-aligned.
- **Feature Extraction:** We organise multimodal lifelog data into a database of minute-long units. We extracted various features from the lifelog data. For example, visual concepts from the image content using automatic and manual annotations as discussed in a previous chapter.
- **Data Anonymisation:** We then apply data anonymisation techniques to anonymize identity of the lifeloggers, bystanders or any other identifiable object. For example, face blurring or other anonymisation procedures such as tattoo blurring, named identity removal, document/screen blurring from wearable camera data. We have discussed a user case study on lifelogger's

privacy concerns to ensure that we adhere to any legal or ethical data governance expectations.

- **Data Fusion:** We then combine multimodal lifelog data to enhance data quality and decision making which is known as data fusion.
- **Lifelog Documents:** After data fusion, we store data into documents of 1 minute length. We have 1440 minutes in a day. The proposed lifelog segmentation algorithms then group the data into a sequential combination of minutes based on basic retrieval units such as moment, activity or event as per the user’s query topic. We term these sequential combination of minutes as lifelog documents.

Nowadays, lifelog annotation and search tasks have been receiving increasing research attention with dedicated workshops and collaborative benchmarking efforts underway [108]. Thus far in lifelog research, the unit of retrieval has been the minute as a time-unit [108] or individual data points. Image, temperature, location [29] and an event are examples. An event is comparatively longer static unit of retrieval and may have overlapped two or more small activities in an hour as introduced in Chapter 3 [4]. We introduce a new unit of retrieval called an activity. Size of an activity depends on the duration of activities of an individual. Activities are more dynamic and produce more accurate and meaningful information for user query requirements. This answers our third research question i.e. what are the various units of retrieval documents for ad hoc search through multimodal lifelog data. This study motivated us to consider lifelog documents for data segmentation tasks.

To address our fourth research question (RQ4) of identifying state-of-the-art indexing time segmentation approach, we did a preliminary experimental study with a small cohort. We collected visual data from 10 participants over a period of one day each i.e. 14132 images. We expected this experiment to identify a baseline segmentation approach [1] for temporal segmentation of visual lifelog data. We present segmentation algorithm based on activity as a basic retrieval unit later in

this chapter. It aims to improve conventional event based temporal segmentation of lifelog data [1]. In this experiment, we used lifelog search challenge (LSC2018) test collection which completely satisfies the proposed principles of lifelog data collection such as continuity, completeness, anonymity and protectiveness as explained in Chapter 3. Following section explains our preliminary experiment and proposed methodology of lifelog event segmentation.

5.3 Event Segmentation of Lifelog Data

The static indexing-time based event segmentation approach temporally segments lifelog data to identify event boundaries. The initial focus of our research is to build a segmentation algorithm for the visual lifelog data. The segmentation approach reported in Chapter 2 implemented similar data analytic techniques to segment lifelog data e.g. implemented in [1]. We noted an opportunity to explore and build lifelog segmentation approach which used multimodal lifelog data along with metadata. The data could be accelerometer readings, temperature condition, infrared lights and physical movements [1] to provide better representation of overall life experiences of an individual.

We investigated [1] segmentation approach as the most appropriate for our experiment where we focused on building visual lifelog data segmentation approaches. We extracted visual features using different visual feature extraction tools such as MPEG 7 descriptors, Caffe framework and MS computer vision. We re-implemented [1] approach as a baseline segmentation approach and proposed two novel approaches for the segmentation of visual lifelog data. We based our first novel approach on extracting visual features using Caffe framework [2] to detect visual concepts that classify an image into 1000 categories of ImageNet classification as described in a previous chapter. Second novel approach is based on extracting categories using Microsoft cognitive services API [3]. These categories were based on parent-child hierarchy where it provides 86 super

categories. It associates each super category with sub categories of various objects in a lifelog image. We present the different segmented user events based on visual features extracted (in Figure 5.2). An example of the types of data streams with their associated daily living small events (see Figure 5.2) is where an individual is peeling carrots in home environment, driving a car and shopping in the supermarket.



Figure 5.2: Example of segmented daily life user events in lifelog dataset using Caffe framework [2] and Microsoft’s cognitive service [3] concept detector tools.

We also show example of the events transition in a particular day from morning such as from “taking blood samples for collecting biometric data in the morning” event to “having breakfast” and then “commuting to work in car” event are shown in Figure 5.3. Major contribution of this preliminary experiment is to identify baseline event segmentation approach for visual lifelog data.

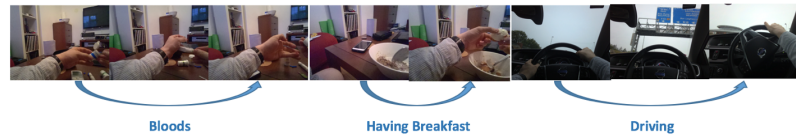


Figure 5.3: Example of signifying the transition between different user events.

We considered several approaches to event segmentation as discussed in Chapter 2. We compared segmentation approaches based on high-level human-understandable semantic visual concepts with basic low-level machine-understandable i.e. computer vision based features such as colour description, colour layout and edge histograms used by Doherty [1]. This shows

the impact of using semantic labels in building content-based lifelog segmentation algorithms. We implemented a methodology to present the process of segmenting visual lifelog data into events as shown in Figure 5.4 based on the methodology proposed in [1].

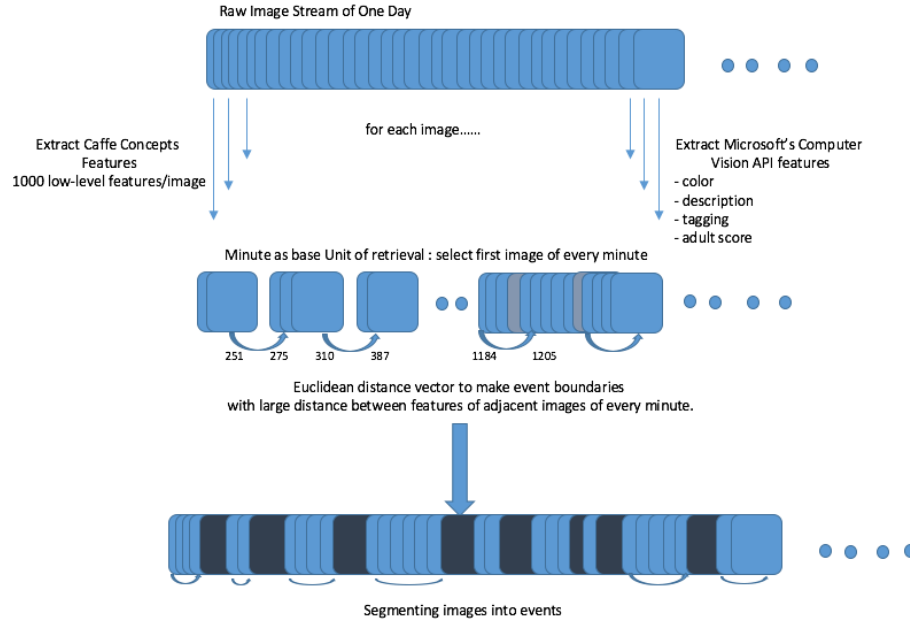


Figure 5.4: Methodology to segment one day visual lifelog data into events at indexing time.

Following the methodology, shown in Figure 5.4. We gathered visual lifelog data from ten participants. We used a wearable camera called autographer. This captures two images per minute. After implementing lifelog data anonymisation techniques as explained in Chapter 3. We cleaned lifelog data by implementing pre-processing techniques as explained in Chapter 4. We extracted visual features using concept detector tools as detailed in Chapter 4. These images are organised into a database of minute-long atomic units by selecting first image of each minute. For each image, we used vector distance method to identify boundaries with high distance between subsequent images. We implemented manual and automatic thresholding techniques to highlight top-ranked event boundaries. We used sliding window approach to aid segmentation algorithm. This refined the similarity between system defined

segmentation points and user defined segmentation points i.e. ground-truth.

5.3.1 Baseline Segmentation Approach to Event Segmentation

To compare against existing approaches as described in Chapter 2, we implemented a baseline segmentation approach based on the work of Doherty et al. [1]. In this approach, authors extracted MPEG-7¹ low-quality visual feature extraction [58] (explained in Chapter 2) from SenseCam images. The authors compared adjacent individual images i.e. non text-tiling approach or blocks of images i.e. text-tiling approach [94] to fine tune lifelog data. They calculated various vector distance methods such as histogram intersection, Euclidean, Manhattan, squared chord, Jeffrey mod KL, Bray Curties, square chi squared, x2 statistics, Kullback Leiber and Canberra to detect the event boundaries. Later, they used automatic threshold determination techniques to highlight top-ranked event boundaries in lifelog data. They used wearable device i.e. SenseCam to record an enormous amount of visual and sensory change in successive images. The neighbouring images proposed as discrete event boundaries even after implementing thresholding methods. Therefore, the authors [1] proposed a post-processing gap approach to identify only one accurate event boundaries and investigated optimum time to ignore the subsequent images proposed as boundaries. A 3 minute gap proved to be the best in [1]. This post-processing boundary gap approach will be explained later in this chapter. The approach to segment visual lifelog data implemented by Byrne [16] as discussed in Chapter 2 is similar to the Doherty [1] approach with minor difference in type of data used. The difference was in light data and human motion data of lifelog images using bluetooth and GPS metadata. Major focus of this research is to build a better visual lifelog event segmentation approach. We followed and extended methodology explained in [1]. We decided to re-implement Doherty [1] approach as our baseline segmentation approach for this research work.

¹<http://image.ntua.gr/iva/files/uploads/tools/vde/documentation.pdf>

5.3.2 Event Segmentation Based on Visual Concepts

It is our conjecture that change in visual objects in the field of view of an individual would cause change of events. This is the first novel approach to event segmentation. We used Caffe framework [2] to detect the objects visible in lifelog image. Caffe framework is BSD-licensed C++ library with Python and MATLAB bindings for training and deploying general-purpose convolutional neural networks and a collection of reference models. We show conceptual overview of segmentation process of observing a change in user event based on Caffe concepts in Figure 5.5.

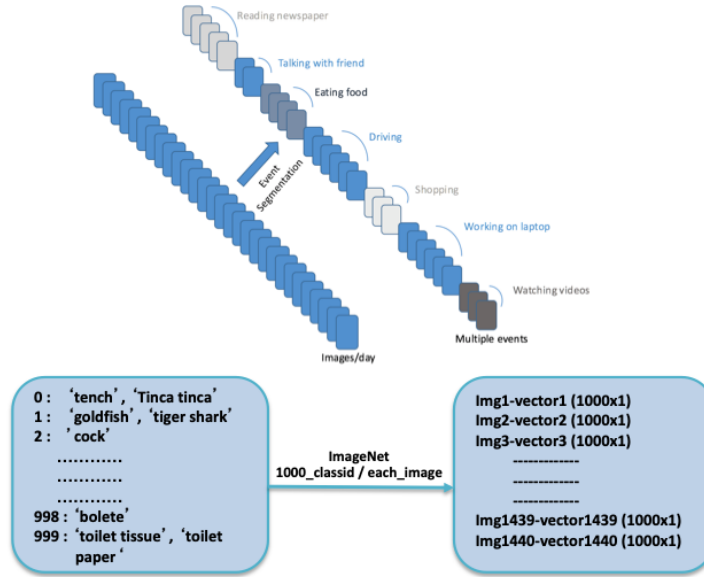


Figure 5.5: Lifelog event segmentation based on Caffe visual concepts of one day data [2].

Caffe is a deep learning framework used with the 1,000 ImageNet dataset of visual concepts [2] and provides feature such as object classification, object detection and learning semantic features using pre-trained network model. Caffe visual concepts form a 1,000 item vector of size 1000 x 1 for each image (see Figure 5.5). The vector stores confidence values of identified objects in image and confidence values for other non-identified objects set to 0. We present the algorithm to segment visual lifelog data using Caffe visual concepts in Figure 5.6.

1. Define a function to read dataset
2. INPUT: dataset attached with visual concepts from Caffe framework
3. Define a empty vector to read all minutes in dataset
4. Scan through all minutes in dataset and select only first image of every minute
5. Define a new vector of size 1000 to hold confidence value of visual concepts for each image
6. Check only for those images which has at least one concept
7. Append concept name and confidence value for each image to the defined vector

Figure 5.6: Algorithm for segmenting visual lifelog data using visual concepts from Caffe framework [2].

5.3.3 Event Segmentation Based on Image Categories

The second novel approach is to utilise high-quality human-understandable semantic labels. Categorisation of images is a source of evidence for the segmentation process in this approach. We have used Microsoft’s cognitive services API [3] in this approach. The API returns categories for each image as described earlier in this Chapter and in Chapter 4. These categories are organised hierarchical categories. There are different categories such as indoor category includes indoor_churchwindow, indoor_door, indoor_room; food category includes food_grilled, food_bread, food_pizza, etc.; outdoor category includes outdoor_mountain, outdoor_city; and building category includes building_arch, Building_brickwall, building_church, building_corner.

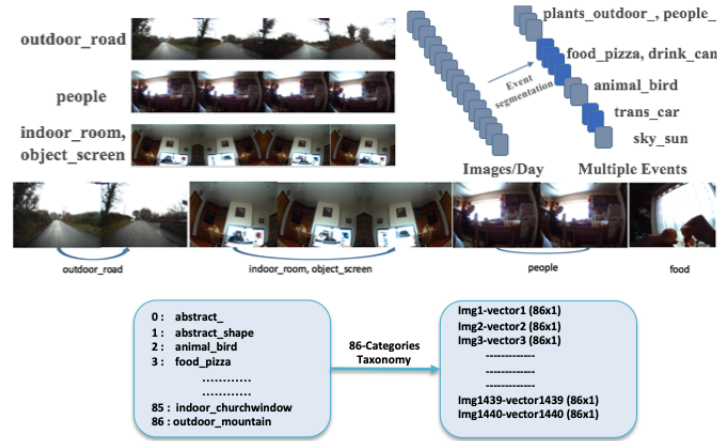


Figure 5.7: Example of segmented lifelog dataset based on categories [3].

This results in one vector of size 86 representing all categories in an image with the associated confidence values. We show it in Figure 5.7. The example identified

various categories in image content and segmented images from a day based on these category taxonomy in Figure 5.7. We also present an implementation segmentation algorithm using image categories as a visual feature from Microsoft cognitive services API.

1. Define a function to read dataset
2. INPUT: dataset attached with visual concepts from Microsoft computer vision API
3. Define a empty vector to read all minutes in dataset
4. Scan through all minutes in dataset and select only first image of every minute
5. Define a new vector of size 86 to hold confidence value of image categories for each image
6. Check only for those images which has at least one category
7. Append category name and confidence value for each image to the defined vector

Figure 5.8: Algorithm for reading lifelog data and visual concepts from Microsoft cognitive services API.

5.3.4 Distance Measure

A key component of event boundary detection is the ability to identify visual or semantic distance between subsequent lifelog images or groups of images. We can declare an event boundary where the distance is above a certain threshold along with other criteria. To calculate the distance, we implemented the Euclidean distance formula on these vectors where each vector represents the confidence score for all visual features extracted from an image. This allows us to identify event boundaries with a high distance and fuse images within the same event with low distance. Therefore, a change in events with high distance such as ‘eating’ and ‘driving’ would be highly distant from each other. So there would an event boundary between these events. Whereas events such as ‘eating’ and ‘cooking’ would have a lower distance and may not trigger an event boundary. This requires selection of automatic thresholds and we describe this in the following section.

5.3.5 Threshold Determination

We need to determine the threshold values to identify top-ranked events boundaries. We implement two automatic thresholding techniques as described in Chapter 4 based on our existing baseline segmentation approach. The first is a parametric

technique i.e. the mean thresholding which takes mean, standard deviation and user parameter k , where k can be of any value between 0.1 to 1.0. Usually, 0.5 is a standardised value for the user parameter k in this technique. The user parameter k can be initialised for the most optimised algorithm performance. We set aside some training data and check variations in accuracy by implementing all values of k to identify top-ranked event boundaries. We found 0.6 to be the best value. The values above 0.6 decrease accuracy to identify top-ranked event boundaries. The second is a non-parametric technique i.e. Kapur thresholding. It is based on entropy where we do not require any user parameter as described in a previous chapter.

5.3.6 Avoiding Over-Segmentation

In lifelogging, there is usually significant visual change in a sequence of camera images because of short-term variation during activities. We needed to avoid over-segmenting the data. Such minor variations are not representative of changes in the actual human activities and we should not segment based on these. To avoid over-segmentation, we chose a minimum five minutes as the smallest duration of a segmented event. We base this event on the best group size of images and identify accurate event boundaries for the human activities.

5.4 Evaluation

We compare both baseline and novel approaches with the event boundaries defined by the lifeloggers themselves which is considered as ground-truth to evaluate similarity measures and the overall system performance. For this evaluation, we developed a new purpose-built visual lifelog dataset gathered by a number of individuals.

5.4.1 Initial Dataset

In this experiment, we collected total 14,132 images, an average of 1,400 images per user from 10 different participants to test both baseline and novel segmentation approaches. All other sensors on the device such as bluetooth and GPS were turned off to optimise battery life. The participants belong to different professions such as expert researchers, early researchers, undergraduate/master students, businessman and house-maker. We asked participants not to change their daily routine due to wearing the camera. To achieve anonymity of lifelog data, we asked participants to adhere to lifelog data governance ethics and guidelines as explained in Chapter 3. Each participant then manually segmented their day into a set of discrete events which is considered as ground-truth. As a result there is a natural variability in the number of events of users regardless of their occupations and human subjective judgements being made. We can point out the numbers of events are in line with what we would have expected. We summarise the detailed information regarding participants and total count of images collected from each user per day in Table 5.1.

Lifelogger/User	Profession	Age group	Duration/day	Images/day	Total ground-truth events/day
1	Researcher	>40	17 hours	1084	46
2	Researcher	>35	12 h 35 m	1792	33
3	Researcher	>35	13 hours	1895	16
4	Researcher	>35	17 hours	1078	31
5	Researcher	>25	11 hours	1064	31
6	Student	>30	11 hours	1399	20
7	Student	>22	11 hours	1494	39
8	Student	>25	14 hours	1336	35
9	Businessman	>55	13 hours	1708	41
10	House-maker	>45	13 hours	1282	19

Table 5.1: The summary information of participants and their count of segmented ground-truth events.

5.4.2 Evaluation Methodology

One challenge when testing event segmentation algorithms is to evaluate segmentation algorithms that produce an accurate segmentation in terms of precision, recall, F1-score and MCC score. But it labels segmentation point to be a

few images before or after the user-defined ground-truth segmentation point. This could be close but not exactly matching human segmentation. It is natural and expected. To solve this problem, we implemented model tuning approaches proposed by Doherty [1]. The post-processing boundary gap approach which ignores the successive event boundaries over an optimum duration of time i.e. 3 minutes in [1] after exactly matching with the ground truth event boundary. It is to be noted that the post-processing boundary gap approach is similar to the sliding window approach to detect nearest possible or exactly matching event boundary with user ground-truth. In sliding window approach, the sliding window is the group of images i.e. a segment of ± 5 images/minutes that runs and slides over the time-series lifelog data in each slide for the whole corpus. In terms of evaluation measurements, we implemented confusion matrix to calculate true positive i.e. where observation is positive and predicted to be positive, false positive i.e. where observation is negative but predicted positive, false negative i.e. where observation is positive but predicted negative and true negative i.e. where observation is negative and predicted to be negative. We reported the results in precision, recall and F1-score as discussed in our baseline approach [1]. Besides these measures, we also utilised Matthew's Correlation Coefficient (MCC) score for our novel lifelog segmentation approaches that makes easier to understand the benefit of using this measure to analyse overall system performance.

We calculated precision in terms of true positives where negative effect of over segmentation is ignored. It is natural that the total count of the event boundaries detected by segmentation algorithm would be high because of the frequent change in visual features as compared to the total count of user-define event boundaries i.e. ground-truth events. Therefore, it is better to over-segment the lifelog data than under-segment to get the high range of relevant event boundaries. Recall represents the fraction of relevant event boundaries that are successfully retrieved by the algorithm and the F1-score represents a harmonic mean of precision and recall. Matthew's Correlation Coefficient (MCC) calculates the correlation

co-efficient between user-defined event boundaries and event boundaries detected by segmentation algorithm between -1 and +1. The segmentation algorithm predicts perfect disagreement where we get the value -1; 0 when there is a random match prediction and 1 when there is a perfect match. MCC scores are based on false positive rate. False positive rate calculates ratio between the total count of false positives i.e. inaccurate detection of event boundary; and the total count of true negatives, i.e. accurate detection of no event boundary. True positive rate calculates the proportion of true positives i.e. correct detection of event boundary. We now present the results of our proposed static indexing-time based event segmentation algorithm.

5.4.3 Results and Discussion

We compare two novel approaches with the baseline segmentation approach as discussed earlier in a comparative study. We arrange all the accuracy scores across the data collected by ten participants over the period of 10 days in this study.

- **Baseline: event segmentation based on MPEG-7 descriptors:** We implemented the segmentation algorithm introduced by Doherty et al. [1]. To identify event boundaries in temporal lifelog data, we use a text-tiling approach i.e. a group of five adjacent images; and a non text-tiling approach, i.e. the adjacent individual images. We used post-processing boundary gap where gap is taken as 3 minutes/images to tune the lifelog segmentation algorithm. In this approach, we ignored the subsequent images that are proposed as false successive event boundaries to the investigated true event boundary, as explained in [1]. Here, we used a confusion matrix to test the model in terms of true positive, true negative, false positive, and false positive values, as described earlier and in Chapter 4. This provides the similarity scores between investigated event boundaries and the ground-truth event boundaries. We then calculated evaluation measures such as precision, recall, F1-score, and MCC value to observe the overall segmentation algorithm performance. It

is clear from Table 5.2 that the text-tiling approach outperforms non-text-tiling approach with the overall improvement of 0.08 in F1-score. We noted that both approaches have positive scores of Matthew’s Correlation Coefficient but non text-tiling approach comparatively performs almost half of text-tiling approach i.e. 0.06. Hence, we found that the text-tiling approach with mean thresholding technique performs better.

Model Tuning Methods	Threshold Value	Precision	Recall	F1-Score	MCC
Text-Tiling based on MPEG-7 Descriptors	mean(K = 0.6)	0.51	0.19	0.28	0.13
	Kapur	0.50	0.18	0.26	0.12
Non Text-Tiling based on MPEG-7 Descriptors	mean(K = 0.6)	0.46	0.13	0.20	0.07
	Kapur	0.47	0.13	0.20	0.07

Table 5.2: Overall Thresholding Performance based on MPEG - 7 Descriptors [1].

- **Event segmentation based on visual concepts:** In this event segmentation approach, we extracted the visual features from visual lifelog data using Caffe framework. We implemented event segmentation methodology as explained in the starting of this section. We implemented the text-tiling approach with mean thresholding technique because the authors proved it to be a better approach in [1]. We tuned segmentation algorithm using post-processing boundary gap approach and introduced a sliding window approach to find the more accurate event boundaries.

Model Tuning Methods	Threshold Value	Precision	Recall	F1-Score	MCC
Sliding window approach	mean(K = 0.6)	0.93	0.35	0.50	0.46
Baseline segmentation approach	mean(K = 0.6)	0.72	0.18	0.29	0.16

Table 5.3: Overall Thresholding Performance based on visual concepts using Caffe framework [2].

As we can observe it from Table 5.3, sliding window approach outperforms the baseline segmentation approach with overall improvement of 0.21 in F1-score. To understand this, the post-processing boundary gap ignores the successive event boundaries over optimum time duration once it found a dedicated event boundary. Beside this, the sliding window approach checks the similarity for truly or nearest matched event boundary in each slide on either sides and selects the nearest event boundary to the user-defined event

boundary. Therefore, we get more accurate event boundaries using sliding window approach and an improvement of 0.30 in MCC score (see Table 5.3).

- **Event segmentation based on image categories:** In this novel approach we segment the visual lifelog data into events based on image categories as described in Chapter 4. We again use text-tiling approach with mean thresholding where user parameter k is 0.6 as discussed earlier. We tune this segmentation approach using post-processing boundary gap approach as explained in baseline segmentation approach [1] and the sliding window approach as described earlier. As we can observe it from Table 5.4, sliding window approach outperforms the baseline segmentation approach with the overall improvement of 0.22 in F1-score and 0.31 in MCC score.

Model Tuning Methods	Threshold Value	Precision	Recall	F1-Score	MCC
Sliding window approach	mean($K = 0.6$)	0.85	0.40	0.54	0.50
Baseline segmentation approach	mean($K = 0.6$)	0.76	0.20	0.32	0.19

Table 5.4: Overall Thresholding Performance based on image categories using Microsoft cognitive service API [3].

The best segmentation approach overall: In Table 5.5, we found that both of our novel event segmentation approaches provide better understanding of visual content available in lifelog data. For example, in two different images where an individual is moving from the office desk to an office kitchen. The visual concepts [2, 3] in the field of view would naturally change from computers to food items. It is likely that a higher-level visual analysis would perform better than one that operates just on lower-level visual features such as edge histogram and colours using the MPEG-7 library as implemented in our baseline segmentation approach [1].

Segmentation Approach	Precision	Recall	F1-Score	MCC
MS visual concepts	0.85	0.40	0.54	0.50
Caffe visual concepts	0.93	0.35	0.50	0.46
MPEG-7 visual concepts	0.51	0.19	0.28	0.13

Table 5.5: Overall thresholding performance based on MPEG - 7 descriptors [1], Caffe visual concepts [2], image categorisation via MS Concepts [3].

It can be observed from Table 5.5 that our novel segmentation approaches

performed better than the baseline segmentation approach [1]. Event segmentation based on Caffe visual concepts [2] improved overall accuracy of segmentation by 0.22 in F1-score and 0.33 in MCC score. Event segmentation based on image categories using Microsoft cognitive service API [3] improved overall accuracy of segmentation by 0.26 in F1-score and 0.37 in MCC score. This analysis justifies that our novel segmentation approaches proved better over the baseline segmentation approach. We also compared both novel segmentation approaches to find the best segmentation approach for the further experiments in this research. It helps us to get a clear vision to proceed for our fourth research question i.e. what is the state-of-the-art static indexing-time based segmentation approach as investigated in Chapter 1. We found that the segmentation approach based on image categories provides better results (see Table 5.5) from the segmentation approach based on visual concepts using Caffe framework with the improvement of 0.04 in both F1-score and MCC score. We can see from Table 5.5 that the segmentation approach based on image categories is the best approach to event segmentation of lifelog data. We conclude that the low-quality features such as MPEG-7 are fast to extract and system friendly e.g. machine-readable which help to increase the speed. The visual concepts are relying on deep neural network approaches such as Caffe visual concepts and image categories from computer vision API help to improve the overall performance of segmentation algorithms. This is our preliminary study to segment lifelog data and the study refers to the static event segmentation approaches that were state-of-the-art approaches as reported in Chapter 2. Later we identified the need of segmenting lifelog data at retrieval time where user can perform variable length activities in a day instead of one fixed duration of segments such as eating an apple in the morning in 3 minutes, eating lunch in 10 minutes or eating dinner in 15 minutes. We reconsider the concept of generating lifelog documents as explained in Chapter 3. We defined activity as a basic retrieval unit. Activity as a retrieval unit could provide a flexible mechanism to segment lifelog data as per user requirement which is one of

the major contributions of this research. Activity is a combination of sequential images/moments whose size depends on the duration of activity of individual and dynamic in nature. Activity would manifest as a fusion of synchronised multimodal lifelog data. We now address our fourth research question i.e. RQ4: what is the state-of-the-art static indexing-time based segmentation approach as investigated in Chapter 1, by building lifelog segmentation algorithm based on activity retrieval unit in the following section.

5.5 Multimodal Lifelog Segmentation Based on Activity Retrieval Unit

Referring to the hierarchical model of units of retrieval explained in Chapter 3, we build a segmentation algorithm based on activity retrieval unit. We have explored previous research focused on an activity-based measure as an indexable unit. For example, Braber used a person's daily food consumption record and environment quality, personal behaviour, biometric data and individual performance in terms of mental and physical activities as a document segmentation model to promote health and wellness [113]. 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 [114]. Such a viewpoint puts the individual at the centre of the segmentation process. It is our belief that the indexing-time segmentation of lifelog data based on activity retrieval unit will be the most appropriate query-agnostic unit for lifelog search and retrieval.

Although the event could be a combination of two or more user activities performed over a longer time as described earlier in this chapter has been considered being the first unit of retrieval for lifelog data. The authors [29] and [12] used event as a basic unit of retrieval in initial Sensecam image viewer tool and in the development of early lifelog search engines. In this section, we test the static indexing-time based lifelog segmentation algorithm based on activity as a

basic unit of retrieval. In activity-based indexing, human activities labelling can range from the broad physical activity measurements such as walking, sitting, running to detailed activities such as looking, speaking and opening a door. For this research, we define a revised set of sixteen lifestyle activities based on but not the same as Kahenmann’s set of the sixteen most enjoyable lifestyle activities [115]. The activities defined in [115] were selected based on personal interview with cohort of users. The dissertation writer had access to for this research and has been previously used as target activity labels in the SenseSeer smartphone-based lifelogging framework [116]. In this study, we define activities after examining the long-term lifelogs of two individuals and represent an achievable segmentation target given the data at our availability. The sixteen lifestyle activities that we have defined are: travelling in car, boat, air-plane, or bus; face to face meetings at work place; using computers or laptops at work; preparing meals in kitchen; eating food with family or friends at home, or at work place; taking care of children; working in the home such as gardening, cleaning, washing clothes, and/or washing utensils; relaxing or watching tv at home; reading paper, socialising at pub or bar, praying in church, shopping in supermarket, playing video games at laptop, engaging in any physical activity, engaging in any creative activity, and other types of activities that are not included in the previous fifteen. The authors detail these sixteen lifestyle activities in [8] and explore knowledge mining and visualisation of lifelogs.

5.5.1 Lifelog Activity Segmentation

We followed the lifelog segmentation methodology as explained in section 5.3 for this experiment. We define minimum document size i.e. moment/minute introduced in Chapter 3 and organise multi-sensor lifelog dataset into minute long segments of 1,440 minutes/day. We implemented pre-processing techniques as explained in Chapter 4 such as data anonymisation and extracted visual concepts. We extracted image categories and image tags using Microsoft cognitive services

API [3]. We calculated Euclidean vector distance method for every moment to identify intra-moment similarity as an implementation of data analytic techniques. We used automatic mean thresholding technique to identify the most effective start and end point of each activity boundary. It is the $\text{mean_threshold} = \text{mean} + K * \text{standard deviation}$ where we decide user parameter $k = 0.6$. At the end, we tuned the segmentation algorithm using the sliding window approach as detailed earlier and post-processing boundary gap approach used by Doherty et al. [1]. In this experiment, we used publicly available lifelog search challenge (LSC2018) test collection as introduced in Chapter 3 to validate the results for the proposed segmentation algorithm. The LSC test collection is a privacy-aware lifelog test collection which facilitates reusable comparative evaluation of information retrieval tasks. We use this dataset to compare static indexing-time based segmentation approaches against dynamic retrieval-time based segmentation approaches later in this dissertation.

5.5.2 Dataset and Groundtruth Generation

In parallel to this research work, the first collaborative benchmarking exercises were being prepared and have been released in recent years under the umbrella of NTCIR [14] and LSC [19]. It has been discussed in detail in Chapter 2. Such datasets give us a sign of queries that a lifelogger will make against their collection for the required information retrieval. LSC2018 lifelog test collection [19] comprises 27 days of multimodal lifelog data with 38,880 moments generated by one active lifelogger. It is a subset of NTCIR-13 lifelog dataset [14]. Various forms of metadata was associated with the images. The metadata could be date/time and high-quality visual features such as categories and tags for each image extracted from [3]. It captures real-world activities of the user.

For this work, we used music listening history data such as song name, artist name and album name; biometric data such as heart rate, galvanic skin response, sleep duration, calorie burn and step count; semantic locations such as home,

work, restaurant); and a manual log of food and drinks. We appended additional metadata to the collection that could be useful for retrieval such as weather, for example rain, fog, sunny, light showers along with temperature conditions using google weather web service. We also appended time of daysuch as early morning, morning, afternoon, evening, night, late night; and manual annotations that will be explained in Chapter 6. The authors have released these additional annotations along with the topics and relevance judgements as an addendum to the LSC2018 collection [78] in July 2019². We show the example of identified image tags and manual annotation of lifelog images where each image is user activity in visual lifelog data along with descriptive metadata in Figure 4.4 in Chapter 4. Lifelogger also generated the ground-truth to provide an unbiased analysis of lifelog data where they refer to generic regular 16 daily user activities. These activities as relaxing, travelling, using any digital screen, face-to-face interaction, praying, playing games on computer, reading paper, preparing meals, socialising, taking care/playing with children, eating, shopping, doing any creative endeavours, physical activity, housework and any other activity as introduced by NTCIR evaluation workshop in lifelog event segmentation sub-task (LEST).

5.5.3 Evaluation Methodology

We tune our proposed segmentation approach against the user ground-truth generated by lifelogger using the sliding window approach and post-processing boundary gap approach as discussed earlier in this chapter. We get the results:

- **Sliding window approach:** When tuning any segmentation algorithm, one of the key challenges is to find a tuning mechanism flexible to minor variations in the exact boundary definitions. For example, a few minutes on either side, as described earlier in this chapter. To resolve this problem, we use sliding window approach around the ground-truth labelled segmentation points. We found that we need to identify the reasonable size of the sliding window to

²LSC2018 Dataset available at: <http://lsc.dcu.ie>

detect accurate user activities. The size of this sliding window could range from 0 i.e. no sliding window allowed up to an arbitrary figure of 16 i.e. upper-bound for reasonable experimentation³). We used several sizes of sliding window w could vary between 1 to 10 to observe the accuracy of detected activity boundaries. We split and slide over entire sequence of images into a set of sliding windows ($\pm [1 \text{ to } 10]$) that allows necessary flexibility in identifying accurate activity boundary in lifelog data. If the system-defined boundary is matched or nearly matched with the ground-truth within a sliding window then it is considered to be accurate. We then fixed that boundary as a segmentation point. Therefore, we have found that five images i.e. window of 5 images which is 10 minutes either side is a reasonable size of sliding window around the ground-truth labelled activity boundaries. This will be validated later in results section.

5.5.4 Results

- **Baseline segmentation approach by Doherty:** The focus of this approach is to identify the most appropriate size of sliding window. We tested the segmentation algorithm over four different sizes of sliding window i.e. window size of 3, 5, 7 and 10 that proved to be a better approach to identify accurate activity boundaries. We summarised the results in Table 5.6. We observe that larger the window size would cause high accuracy in retrieving user activities but would reduce overall system performance in terms of recall value. The sliding window with ± 10 subsequent images i.e. 20 minutes performs best. In this approach, we consider MCC value for measuring the accuracy of our proposed algorithm. We get overall performance of the segmentation algorithm with 0.52 as F1-score and 0.22 as MCC score. By using post-processing boundary gap model tuning approach

³Given a 16 hour day, with one image per minute and 30 events identified per day (all reasonable assumptions), then an evaluation with a 16 minute boundary would tend not to penalise random segmentation algorithms.

we found that the proposed segmentation algorithm is successfully identifying most of the activity boundaries with high scores of precision but with poor recall values which are not preferable to build an ad hoc search engines at the later stages in dynamic lifelog segmentation approach. Therefore, we infer that we need a segmentation approach that balances the precision and recall scores to build better segmentation approaches.

Table 5.6: Various evaluation approaches implemented to calculate the accuracy of static (indexing-time) activity boundary detection and segmentation using image categorisation from MS visual concepts [3] and identify the best evaluation approach.

Lifelog Activity Segmentation Approach	Frame Size	Precision	Recall	F1-Score	FPR	TPR	MCC
Baseline segmentation approach by Doherty	3	0.62	0.11	0.20	0.08	0.11	0.04
	5	0.86	0.20	0.34	0.07	0.20	0.17
	7	0.88	0.28	0.42	0.12	0.28	0.16
	10	0.93	0.36	0.52	0.11	0.36	0.22
Activity based segmentation approach	1	0.74	0.11	0.20	0.30	0.74	0.21
	2	0.90	0.20	0.32	0.28	0.90	0.34
	3	0.95	0.27	0.42	0.28	0.95	0.42
	4	0.98	0.32	0.48	0.28	0.98	0.46
	5	0.98	0.38	0.54	0.28	0.98	0.52
	6	0.98	0.43	0.59	0.27	0.98	0.56
	7	0.99	0.47	0.64	0.26	0.99	0.58
	8	1	0.50	0.66	0.26	1	0.60
	9	1	0.53	0.69	0.26	1	0.62
	10	1	0.55	0.70	0.27	1	0.63
Overall Best Segmentation Approach	Frame Size	Precision	Recall	F1-Score	FPR	TPR	MCC
Baseline segmentation approach by Doherty	5	0.86	0.20	0.34	0.07	0.20	0.17
	10	0.93	0.36	0.52	0.11	0.36	0.22
Activity based segmentation approach	5	0.98	0.38	0.54	0.28	0.98	0.52
	10	1	0.55	0.70	0.27	1	0.63

- **Activity based segmentation approach:** We initially focus on conventional measurement of evaluation such as F1-score where we completely ignore the true negatives i.e. TN is taken as 0 where user activities neither available in user ground-truth nor detected by system. We get 0 value for true negatives and values for true positive rate, false positive rate and MCC value gets unbalanced. This justifies the need of considering true negatives in our experiment to get measure system performance. In addition, we observed that recall value is almost half of the precision value as seen in Table 5.6). This is because of the nature of indexing-time segmentation where all activity boundaries are pre-defined and fixed. Activity boundaries defined by users are around 25-30 per day as compared

to activity boundaries defined by our proposed segmentation algorithm which is around 80-90 per day. This results in an increase in precision where almost all activity boundaries get detected accurately while evaluating segmentation algorithm but reduces the recall value because of missing many inaccurate activity boundaries.

To tune the models, we need to identify the most appropriate size of the window in sliding window approach. It is our belief that bigger the window size has to provide a higher score as we can observe it from Table 5.6. We get the highest scores of 1 precision, 0.55 recall, 0.70 F1-score and 0.63 MCC score with window size 10. As we discussed earlier, choosing window size 10 broadens the possibility to identify accurate user activities as per user requirements. We also observe that the accuracy measure gets constant at particular window size where precision score reaches at 1 and trade-off between precision-recall value becomes constant for the further sizes of sliding window. We found that the sliding window comprising five images i.e. 5 images = 10 minutes is the most appropriate size of the sliding window for evaluating both proposed activity segmentation algorithm because the trade-off between precision and recall is more consistent to get almost similar scores for both precision and recall.

Overall best segmentation approach: As we can observe from Table 5.6, we get the highest scores of precision (1), recall (0.55), F1-score (0.70), and MCC score (0.63) with the size of the sliding window ($w = 10$) using the sliding window approach. We improved the MCC score where window size = 10 i.e. 0.63 by (0.41) when compared to baseline segmentation approach (0.22) proposed by Doherty et al. [1]. Although in case where we more care about F1-score, we observe improvement in F1-score when window size is 10 i.e. 0.70 by (0.18) when compared to baseline segmentation approach (0.52) [1].

5.6 Chapter Summary

In this chapter, we presented a preliminary study by introducing two novel approaches for static indexing-time based event segmentation of visual lifelog data. We based the novel approaches to high-quality visual feature analysis [2, 3]. To place our research in past state-of-the-art, we defined a baseline segmentation approach to segment visual lifelog data into retrievable events based on the work of Doherty et al. [1]. We compared our two novel approaches to the baseline segmentation approach in an experimental setting with the lifelog data of ten users. We showed that our novel approaches perform significantly better than [1]. This suggests that there is a significant scope for enhancing performance of event-segmentation algorithms on lifelog data and that this is far from a solved problem. We note that these are preliminary approaches and are subject to optimisation and enhancements. The fact that we only analysed visual content also limited this work. We reconsidered the process of generating lifelog documents to explore the other unit of retrieval. We then defined activity as the most appropriate unit of retrieval in this dissertation. Following this study, we evaluated the baseline segmentation algorithm using publicly available lifelog search challenge (LSC2018) test collection. We went on to explore optimal window size while evaluating multimodal lifelog test collection [10] and resolved our fourth research question i.e. identifying the state-of-the-art indexing time segmentation approach. In the next chapter, we will extend this work by developing dynamic query-time based segmentation algorithms so we can compare our proposed approaches with the full spectrum of alternative approaches (i.e. indexing-time v/s query-time). This would address our fifth research question i.e. Do a dynamic query-time based document segmentation approach outperforms the conventional indexing-time based event segmentation approach for ad hoc lifelog retrieval and would justify the hypothesis of this dissertation.

Chapter 6

Dynamic Segmentation of Multimodal Lifelog Data

6.1 Introduction

In a previous chapter, we discussed our static indexing-time based lifelog segmentation approach and validated activity as an appropriate unit of lifelog data retrieval in this research. Activity is defined as a continuous sequential state of an individual in terms of the environment or stimuli that represent a combination of sequential moments whose size depends on the activities of the individual [80]. In this chapter, we focus on developing dynamic query-time based segmentation algorithms where basic unit of retrieval is an activity. We have presented static and dynamic segmentation approaches in Figure 6.1. It is to be noted that subjective human judgement of where an in-vivo activity ends and the next one starts is unlikely to be clearly represented in sensor data stream. Problem with static segmentation approach is that the user define all activities at indexing time and once the segmentation is done then it cannot be changed at retrieval time. We have shown it in Figure 6.1 as the static approach. Therefore, we propose an alternative segmentation approach called dynamic segmentation which is performed at retrieval time and attempts to define a lifelog document dynamically i.e. activity that best matches the exact user information needs at retrieval time. This would enable user to change the topic of interest as per requirement through ad hoc query modification. Query-time segmentation algorithms would benefit

community to build optimised ad hoc search engines to retrieve meaningful information from multimodal lifelogs.

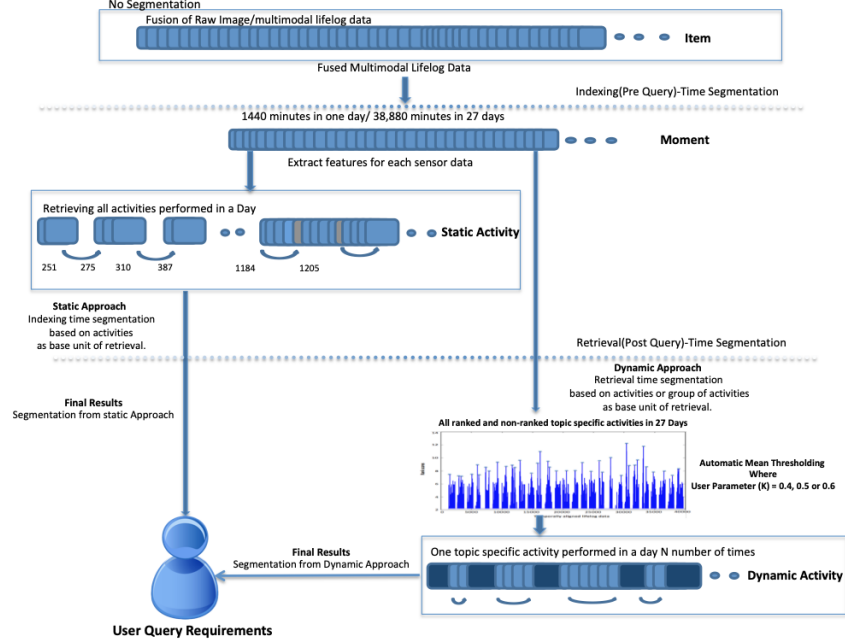


Figure 6.1: Working on activity based dynamic segmentation of lifelog data.

This chapter is outlined as: Section 6.2 presents proposed dynamic segmentation approach based on activity as basic unit of retrieval. We engaged in continuous experimental studies to improve accuracy of segmentation algorithms by including metadata as described in Chapter 2 and considered impact of manual annotations for improving retrieval. We present a comparative evaluation using Lifelog Search Challenge i.e. LSC2018 [78] dataset as described in a previous chapter to identify the best approach to lifelog segmentation by comparing both static and dynamic approach in section 6.3. We show the best performing lifelog segmentation algorithm which addresses the hypothesis of this research work and resolves our last research question i.e. the dynamic query-time based document segmentation approach outperforms the conventional static indexing-time based lifelog segmentation approach for ad hoc lifelog retrieval.

6.2 Dynamic Activity based Segmentation of Multimodal Lifelog Data

We propose a dynamic segmentation approach to segment multimodal lifelog data based on activity as a basic unit of retrieval. It is our inference that dynamic segmentation provides fast and reliable ad hoc retrieval from lifelog storage in terms of activity as a lifelog document. This facilitate users to retrieve information efficiently. It is motivated by previous studies [67–69, 78] and reported in Chapter 2. First, we explain the methodology that detects topic-specific activity boundaries at query-time as per user information needs:

1. Collect multi-sensor lifelog data and define the smallest indexable unit of retrieval e.g. one minute called a moment [11].
2. Extract automatic visual concepts using a standard concept detector tool. Also collect multi-sensor data such as activity logs, music listening history, biometric data, semantic locations. Manually annotate the data such as eating, relaxing, watching television, using a laptop in an office environment. We will detail this in section 6.2.1. In addition, we need to collect metadata such as date or time feature, age, height, weight, manual diet log and weather status for each moment or image in lifelog data and index the corpus in a text index.
3. Fuse the multi-sensor lifelog data into minute long segments i.e. moments for indexing lifelog data into a database of minute-long units. We have 38880 moments spanning over 27 days in LSC2018 lifelog test collection.
4. Generate a temporally ordered query-similarity vector for all moments in the collection by implementing text ranking algorithms as introduced in Chapter 2 upon receiving a user query. Implement a standard vector distance method such as Euclidean distance to calculate intra-moment similarity after vector formalisation.

5. Fuse similar scored sequential moments in one cluster. We declare a boundary if the similarity between moments surpasses a pre-defined threshold using Euclidean distance method and automatic mean thresholding method as discussed in a previous chapter.

In the next section, we present a detailed implementation of ranking algorithm and data analytic techniques such as Euclidean distance method and mean thresholding method.

6.2.1 IR Experimental Methodology

TF-IDF [73] and Okapi BM25 [63] are two basic text ranking algorithms as discussed in Chapter 2. These algorithms rank the best matching documents according to their relevance to a search query. We implemented both ranking algorithms and found a minor difference between the two which is most likely because of the relatively small size of the collection i.e. 38880 moments. We chose BM25 because of the greater potential for parameter optimisation for our experiments. In addition, we included ranking factors such as image optimisation. We used logical operators, boolean operators, Ngrams and stemming for the search queries. An example of retrieved top-ranked query-focused activities for searching moments such as images where user is using digital screens is shown in Figure 6.2:

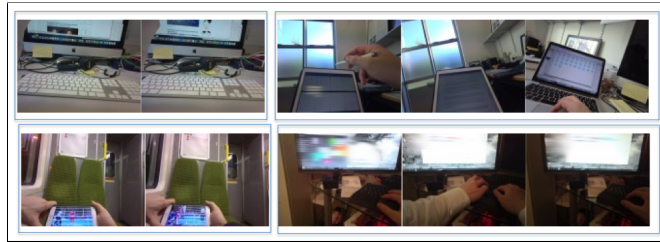


Figure 6.2: Example of top ranked query-focused retrieved activities i.e. four different visual shots of detection and segmentation typically belong to one specific activity but four different day environments e.g. Query: find where the user is using digital screens.

The BM25 algorithm helps to generate a query-similarity score vector for all ranked moments which was specific to the user query topic. We associate these

ranked documents with normalised score values in between 0 to 1. The non-ranked documents usually represent empty score values as illustrated in Figure 6.3. User searching with query “*Find the time where I am using digital screens*” is one such example. It is natural that user must have used digital screens several times in 27 days . Therefore, our algorithm is able to retrieve top-ranked documents for this query as shown in Figure 6.3.

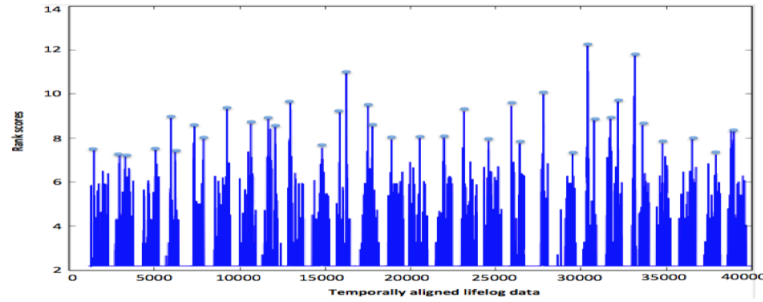


Figure 6.3: Process of identifying effective top-ranked activity boundaries in LSC Dataset.

Applying Euclidean vector distance method between these adjacent documents to generate document boundaries (see formula 6.1). A high distance identified document boundaries as opposed to related activities within a document. This facilitated ranking of top documents via a length-normalised overall score for the documents.

$$Euclidean_Distance(V) = \sqrt{\sum_{n=1}^{38880} (v_2 - v_1)^2} \quad (6.1)$$

To highlight the top-ranked activity boundaries that reflect user information needs, we implemented a document cut-off threshold based on an automatic mean thresholding technique. Refer to Figure 6.3. Mean thresholding technique uses mean, standard deviation and user parameter values where k can be between 0.1 to 1.0. We defined value of K as 0.6 in training phase which provided the most optimised system performance as explained in the previous chapter. We show the example of highlighted top-ranked activity boundaries in Figure 6.3 i.e. circled

points on the bars and visually represented in Figure 6.4.

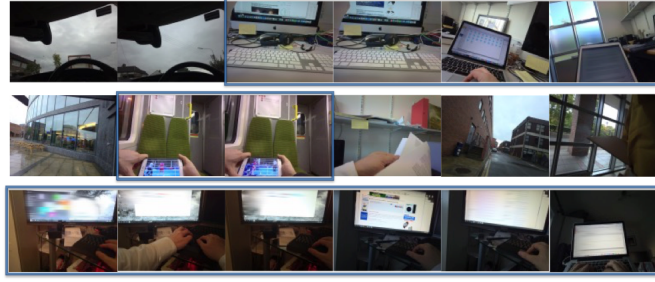


Figure 6.4: Visual representation of identified human activity such as find the time where an individual is using digital screens in LSC dataset.

6.2.2 User Query Topics

Dynamic segmentation approach assumes that there is user information needs to seed the segmentation process. We used existing lifelog dataset from NTCIR13-Lifelog [8] evaluation workshop to generate such information needs. We utilised fifteen broad topics that represent user’s daily life activities as introduced in lifelog event segmentation task i.e. LEST, see appendix A.1. LEST is a sub-task of NTCIR13-Lifelog evaluation workshop as described in Chapter 2. These fifteen user query topics are praying, playing games on laptop, using computer, travelling, face-to-face interaction, physical activity, home environment, creative endeavours, shopping in any shopping mall or supermarket, socialising outside home/workplace, playing with children, eating meals, cooking, relaxing, reading or writing on paper in physical medium. These topics provide full coverage of the expected regular life activities of an individual and formed basis of the information needs. However, we observed that the LEST user query topics were limited to these proposed topics rather than free text.

Therefore, a connection was set up between visual concepts and LEST user query topics during lifelog data pre-processing to integrate ad hoc search process with segmentation process. For example, a user wishes to search for a topic “Travelling”. As we can see from the appendix A.1, we have the “event” as search

query type, “the individual is travelling” as topic description, and “provided a detailed guideline for the user to customise his/her search” in narrative to explain the user query topic. For example, we searched for the query “find the time where the user is travelling in bus, train, boat, bicycle and airplane?” and received top-ranked relevant documents. We utilised LSC2018 corpus for the LEST user query topics here. Thus, we clustered top-ranked relevant documents to integrate the query process into the segmentation process. We evaluated the results of the segmentation approaches in terms of F1-score as measured in our baseline segmentation approach [1]. So we can help research community to understand critical and comparative analysis of proposed segmentation approaches and the baseline segmentation approach. We also calculated MCC score to measure and maximise the overall accuracy of the lifelog segmentation model.

6.2.3 Evaluation Methodology

In this initial experiment, we followed experimental configuration as explained in Chapter 5. We used sliding window approach with sliding window size as 5. We used precision, recall, F1-score and MCC score to evaluate the model. We compare static segmentation algorithm as discussed in the previous Chapter against our proposed dynamic query-time based segmentation.

6.2.3.1 Results and Discussion

Following experimental configuration explained in Chapter 5, we searched for each user query topic as listed in Table 6.1. We achieved the highest score of precision (0.92), recall (0.99), F1-score (0.95) and MCC score (0.93) for praying activity with search criteria *Query keywords = Church, indoor*. More than half of the activities performed better in the dynamic segmentation. Some activities performed poorly, because those activities were driven by low recall values. We believe that it could be a function of thresholding values that we used when selecting cut-off of top-ranked documents in ranked list. We also note that there is

a significant potential for activity crossover for poorly performed user query topics such as eating, shopping and socialising activities. An activity is not highly ranked because of absence of expected visual concepts, or where the activity is easily confused with an alternative activity such as eating vs socialising. Consider a scenario in which user is eating food and engaged in conversation with family members or friends in a restaurant or outside home environment. The search criteria for eating activity is open to any location that includes home, work place or restaurant. So we searched for *keywords: table, plate, food, pasta, pizza, sandwich, home, restaurant, workplace*. The search criteria for socialising activity includes face-to-face conversation with other people in a non-work and non-home environment. So we searched for the *keywords: restaurant, people, group, outdoor*. Therefore, retrieved results may overlap the eating activity while having conversation with people in the restaurant which makes difficult to identify the activity uniquely with segmentation algorithm.

Table 6.1: Dynamic event segmentation based on activity or group of activities as basic unit of retrieval (Dynamic Approach in Figure 6.1).

LEST Topics Query	Precision	Recall	F1-score	MCC score
Praying	0.92	0.99	0.95	0.93
Gaming	0.98	0.70	0.82	0.87
Using computer	0.72	0.90	0.80	0.85
Reading paper	0.99	0.66	0.79	0.81
Face to face interaction	0.99	0.64	0.78	0.77
Travelling	0.71	0.82	0.76	0.71
Cooking	0.74	0.78	0.76	0.67
Relaxing	0.99	0.60	0.75	0.61
Creative endeavours	0.96	0.61	0.74	0.54
Socialising	0.71	0.77	0.74	0.46
Physical activity	0.75	0.72	0.73	0.42
Home	0.80	0.66	0.72	0.33
Children	0.79	0.66	0.72	0.27
Eating	0.96	0.54	0.69	0.21
Shopping	0.95	0.51	0.66	0.07
Overall System Performance	0.86	0.70	0.77	0.57

From Table 6.1, average system performance of dynamic activity based segmentation in terms of 0.77 F1-score and 0.57 MCC score outperforms static segmentation approach where average system performance is less in terms of 0.54 F1-score and 0.52 MCC score as shown in Table 5.6 in Chapter 5. However, we still need to improve the results because we selected very broad topics for this

initial experiment from the NTCIR 13-Lifelog LEST i.e. event segmentation task. A typical ad hoc user query has a narrower focus. For example, the time I was taking a photo of my family by the sea after a picnic or the time I was talking with a person who has a ponytail will be addressed in forthcoming experiments. This type of user query topics are much more suitable for a dynamic segmentation approach rather than broad user query topics. We extended our initial experiment by introducing manual annotations. We also introduced 24 new broad and narrow focused user query topics that intended to provide fair and repeatable comparison.

6.3 Considering Manual Annotations

The basic premise of this work is on our conjecture that manual annotations should produce documents that better fit a user's information need. We extend previous work by proposing three segmentation approaches based on high quality automatic, manual and fusion annotation as shown in Figure 6.5. In this experiment, we show the impact of using manual annotations in the dynamic query-time based segmentation algorithms.

6.3.1 Automatic Annotation Based Segmentation

We show automatic annotation based segmentation approach in Figure 6.5. For **automatic annotation** based segmentation approach, we extracted visual concepts and descriptions generated by a high quality visual concept detector [3]. This concept detector operates over an image from lifelog data and produces high-quality visual features; semantic metadata such as background colour, foreground colour, dominant colour; description of an image such as caption; categories based on 86-hierarchical categories taxonomy for each image; and image tags as discussed in Chapter 4. For this experiment, we used tags relevant to the content of the particular image of lifelog data that include objects such as laptop, television, chair, table, knife; living beings such as person, child, man, woman;

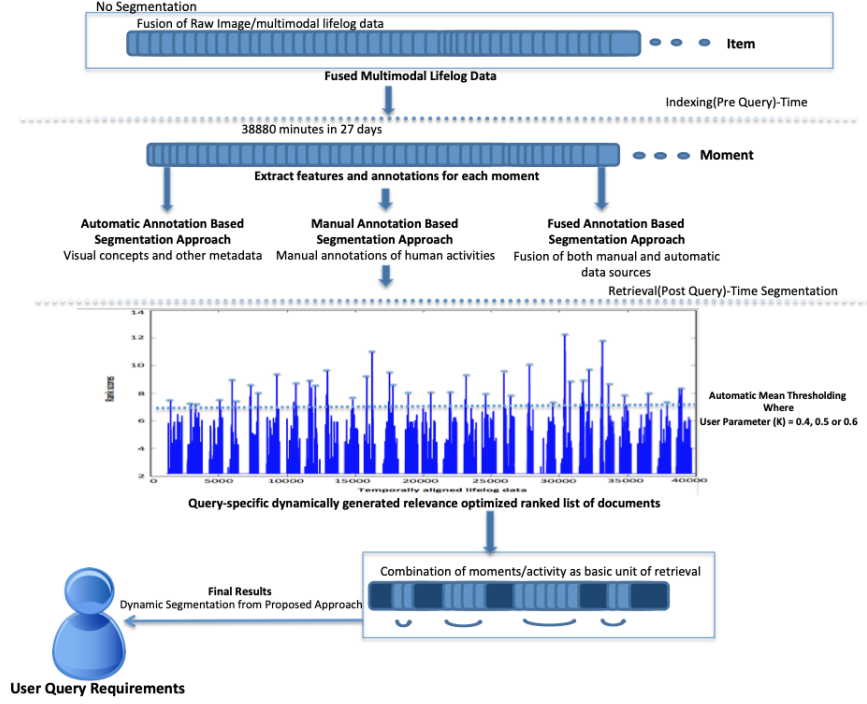


Figure 6.5: Considering manual annotations in activity based dynamic segmentation of multimodal lifelog data.

scenery such as sky, lawn, green, residential, sea; or actions such as working, standing, sidewalk. An example image with high-quality visual concepts and descriptions is shown in Figure 6.6.

6.3.2 Manual Annotation Based Segmentation

The **manual annotation** based segmentation approach was amended with annotations from an ontology of 24 real-world life activities (see appendix A.2 and A.3). We introduced this ontology by an in-vivo process between the lifelog data owner and the researcher. It considers broad to narrow focused information needs of lifelog data owner. A manual review of moments labelled 24 activities by one expert researcher over a number of days. The ontology was single-level and included activity concepts. We show manual annotation based segmentation approach in Figure 6.5. The 24 manually annotated user activities are such as commuting to work, travelling, preparing meals, eating/drinking, taking care of

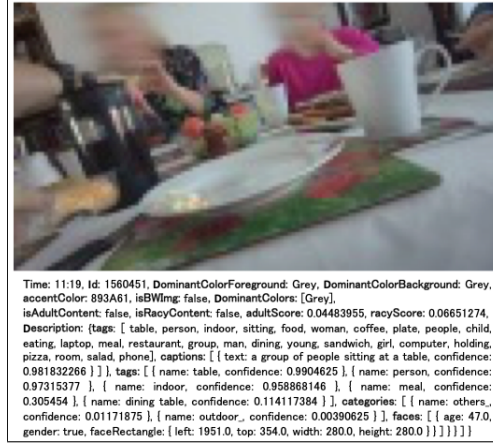


Figure 6.6: Examples of automatic high-quality visual concepts from concept detector using Microsoft cognitive services API [3].

children, praying, socialising/casual conversation, reading, gardening, shopping, work meetings, watching TV, playing computer games, using laptop/desktop computer, using mobile/tablet, any physical activity, sleeping, relaxing, organising things, packing, cleaning, hygiene and make-up activity, writing on paper and searching/information seeking. We show an example image with manual annotations in Figure 6.7.



Figure 6.7: Example of manual annotations based on 24 real-world life activities.

6.3.3 Fused Annotation Based Segmentation

For **fusion annotation** approach, we used fusion of both automatic and manual annotations. We amended these annotations to the metadata for indexing and provided a fair comparison with proposed segmentation approaches as discussed

earlier. We show fused annotation based segmentation approach in Figure 6.5. Once we had generated lifelog corpus based on these three approaches just described, we implemented the method explained in section 6.2. In addition, we fine-tuned segmentation algorithm using sliding window approach as discussed in Chapter 5 [1, 80]. This provide a repeatable comparison among the new proposed approaches and we will discuss it later in this section.

6.3.4 Developing New User Query Topics

The LSC2018 dataset includes 18 different user query topics. The topics were specifically introduced to support interactive experimentation which was out of the scope for our research. We developed 24 new user query topics to reflect wide range of human activities. We expect these topics to be used with lifelog datasets based on the proposals in [17]. This motivates development of different queries for lifelog retrieval systems. We developed 12 broad focus query topics that support lifelog reminiscence/reflection and 12 narrow-focus query topics that reflect conventional retrieval needs as listed in Table 6.2. Examples of broad focus user query topics are such as shopping, reading, working, driving and cleaning. Examples of narrow focus user query topics are such as waiting for a train, packing a suitcase, walking on a lovely day, eating an apple, brainstorming, having a talk with a person who has a ponytail. We have manually generated complete relevance judgements for each query topic. We have also released with the topics in the dataset addendum.

6.3.5 Evaluation Methodology (Precision Focused)

The primary motive of this experiment was to evaluate impact of annotation quality on performance of a novel dynamic query-time based lifelog segmentation algorithm. To do this, we need to test quality of lifelog segmentation given the three different annotation methodologies as described earlier. Defining activity boundaries in lifelog data is inherently subjective and evaluation methodology

Table 6.2: List of 12 broad focused and 12 narrow focused user queries (topic titles only).

Query Topics	
Broad Type Queries	Narrow Type Queries
Shopping	Saturday Morning Coffee
Reading	Walking to the Airplane
Cleaning	Waiting for a Train
Resting	Cutting the Grass
Driving	Walking at Work
Flying	Eating an Apple
Working	Writing on Paper
Socialising	Brainstorm
In a Meeting	Fruit Bowl
Cooking at Home	DIY Store
Watching TV at Home	Packing
Walking on a Lovely Day	Ponytail

used should be robust to minor variations in boundary definitions. Therefore, we employed a sliding window ± 5 minutes i.e. images/moments, which we believe depends upon user query topics and data. For example, if we chose ± 5 minutes as the best size of a sliding window in the previous Chapter. This provides a necessary flexibility in measurement process as explained in Chapter 5. If a system-defined activity boundary is within ± 5 minutes (images/moments) of the human judgement then it is considered accurate. This approach was taken for segmentation evaluation in [1].

We temporally clustered top-ranked moments into candidate activities and a mean thresholding method was used to select only highly ranked activities for the evaluation. Sliding window was used to identify how accurately it detected boundary and any boundary found that was greater than ± 5 minutes (images/moments) was considered being incorrect. This methodology provides precision values (see formula 6.2) to be calculated in terms of true positives. In this evaluation, the negative effect of over segmentation i.e. false positives were ignored as is the recall of low-ranked events relevant to the topic. The reason for this was to focus on finding small number of relevant documents in most cases for a user query topic.

$$\boxed{Precision(accuracy) = \frac{TruePositive}{TruePositive + FalsePositive}} \quad (6.2)$$

6.3.6 Results and Discussion

In this experiment, we proposed and compared three different segmentation approaches i.e. automatic annotation-based segmentation, manual annotation based segmentation and fused annotation based segmentation. We present the importance of manual annotations in dynamic event segmentation approach by developing 12 broad and 12 narrow focused user query topics. Now, we explain the results of these three different approaches of segmentation:

- **Automatic annotation based segmentation:** A mean precision value 0.88 for broad topics and 0.86 for narrow topics in terms of overall accuracy of the segmentation approach based on the image tags description was calculated. We summarised results in Table 6.4. We found little difference between broad and narrow topics when retrieval was based on the automatic indexing process. There was a clear mismatch between the visual concepts extracted by the visual concept detector tool and topics defined by lifelogger. The narrow-focus topics such as eating an apple (0.67), brainstorming (0.62) and ponytail (0.45) proved to be comparatively difficult. It can be observed from Table 6.3. We identified some valid reasons for this. For example, we have an image where an individual is eating an apple. The concept detector tool labels that image as this image contains food item, instead of providing detailed information about food item such as an apple or a fruit. Another example is for the topic where an individual is engaged in work meeting. Visual concept detector labels image such as indoor, person, board which were semantically distant from the topic. In a narrow-focused topic such as talking to a man with a ponytail was more detailed than related concepts such as people/person. Ultimately, performance of automatic segmentation approach could only be increased if

either there were more concepts or the topics were somehow translated from the human semantic level into the system level. This is a variation of the well-known ‘semantic gap’ from multimedia retrieval.

Table 6.3: Performance (i.e. precision) of the dynamic lifelog segmentation based on manual annotations and automatic visual concepts.

Broad Topic User Queries				Narrow Topic User Queries			
Topics	Auto Annotation	Manual Annotation	Fusion	Topics	Auto Annotation	Manual Annotation	Fusion
Shopping	0.81	0.99	0.99	Saturday morning coffee	0.99	1	1
Reading	0.90	0.99	0.99	Writing on paper	0.56	0.99	0.99
Cleaning	0.99	0.99	0.99	Waiting for a train	0.98	1	0.99
Socialising	0.98	0.99	0.99	Cutting the Grass	0.97	0.99	0.99
Driving	0.97	0.99	0.99	Packing	0.94	0.96	0.98
In a Meeting	0.96	0.99	0.98	Walking at Work	0.93	0.94	0.96
Watching TV at home	0.91	0.99	0.97	Walking to the Airplane	0.86	0.99	0.88
Flying	0.81	0.96	0.96	DIY Store	0.78	0.93	0.87
Cooking at home	0.88	0.87	0.96	Eating an apple	0.67	1	0.86
Working	0.80	0.97	0.96	Fruit	0.67	0.76	0.72
Walking on a lovely day	0.75	0.86	0.89	Brainstorm	0.62	0.69	0.72
Resting	0.74	0.73	0.87	Ponytail	0.45	0.56	0.45
Average System Performance	0.88	0.94	0.96	Average System Performance	0.86	0.90	0.88

- Manual annotation based segmentation:** Manual annotations are more reflective of human semantic descriptions and are more related to the human-generated information needs. The precision was significantly higher at 0.94 for broad query topics because of a clear semantic match with the human annotations which should not come as a surprise. However, there was descriptive difference between scores for certain broad topics such as cooking at home (0.87), walking on a lovely day (0.86), resting (0.73); and for narrow topics such as looking at bowl full of fruits (0.76), brainstorming (0.69), and talking with a person who has ponytail (0.56) the precision levels were significantly reduced (see Table 6.3). It is worth noting that these are mostly narrow-focus topics. The average precision of this approach is 0.94 for broad-focused topics and 0.90 for narrow-focused topics. This makes an improvement over the automatic annotation approach, as summarised in Table 6.4. It leads us to conclude that accuracy of annotations is a significant factor in the overall performance of a segmentation algorithm.
- Fused annotation based segmentation:** Combining both automatic and manual annotations provided best results with precision score of 0.96 for broad-focused topics. This achieved an increment of 0.08 as compared to automatic annotations based segmentation. It also reported slightly worse

results 0.88 i.e. an increment of 0.02 as compared to automatic annotations based segmentation for narrow-focused topics as summarised in Table 6.4. It is interesting to note that manual annotations improved the results for many of the poorly performing topics when compared to the automatic approach such as writing on paper, eating an apple, shopping, flying in an aeroplane, working on a laptop and walking on lovely day. We show the results in Table 6.3. This reinforces our hypothesis that enhancing the performance of the annotation engines has a significant impact of segmentation algorithm.

Table 6.4: Comparing system performance (i.e. average precision) of query-specific dynamic event segmentation based on three different approaches to annotation.

Query Topics	Auto Annotation	Manual Annotation	Fusion	Max Improvement
Broad Type Queries	0.88	0.94	0.96	0.08
Narrow Type Queries	0.86	0.90	0.88	0.02

Therefore, we found that manual annotations of human activities increases precision score of the proposed dynamic query-time based segmentation approaches for both broad and narrow focus user query topics. Although, we find a low increase (0.02) in narrow-focused topics but it cannot be claimed to be significant. We hypothesise that this is because human annotations are both higher quality, accurate and more semantically meaningful than the automatic approaches. We also point out that the positive effect of manual annotations is likely to decrease as more accurate and semantically meaningful automatic approaches to annotation of lifelog data become available in the coming years.

It is to be noted that the study was carried out using the dynamic (query-time) segmentation algorithm, as discussed in section 6.2 and evaluated using the reusable LSC2018 lifelog test collection, as described in Chapter 5. We can suggest that all lifelogs can benefit from a human annotation of life activities. We realise that this is unlikely to occur in most real-world scenarios due to the human overhead of making manual annotations. We propose that we require more detailed automatic annotations along with a research focus on developing activity-based annotations similar to the annotations generated by humans. We now focus to identify the best

segmentation approach by comparing the static indexing-time based segmentation approach, as discussed in Chapter 5 and the dynamic query-time based segmentation approach, as discussed in this chapter in the following section.

6.4 Identifying the Best Segmentation Approach

To identify the best segmentation approach in the static or the dynamic, we implemented the methodologies explained in section 6.2 and used the sliding window approach, as discussed in Chapter 5 to tune the proposed segmentation algorithms. We evaluated the segmentation algorithm by using precision, recall, F1-score, and MCC score and utilised LSC2018 dataset. We explain this evaluation algorithm in Figure 6.8.

1. Define a function for sliding window approach
2. INPUT: list of user groundtruths and system defined activity boundaries
3. Iterate the steps 3 to 5 over the frame size in the range of 1 to 10
3. Create a relation matrix for user groundtruth activity boundaries
4. Create a relation matrix for system define activity boundaries
5. Measure the similarity between these two relation matrices
6. Use confusion matrix metrics (true positive, true negative, false positive, false negative) to calculate
 | precision, recall, f1 score, system accuracy, false positive rate and true positive rate

Figure 6.8: The sliding window approach for the evaluation of activity detection and segmentation algorithm.

In this experiment, we focused on identifying the most appropriate size of the sliding window, as discussed in Chapter 5. We utilised different sizes of sliding window w (where $w = 1$ to 10 i.e. $\pm[1$ to $10]$ minutes, see Figure 6.8) to observe the accuracy of detected activity boundaries. We found that the frame comprising $\pm/5$ minutes is the most appropriate size of the sliding window for tuning both the static and the dynamic segmentation algorithms based on activity retrieval unit. The reason for this was the trade-off between precision and recall at $w = 5$, i.e. 10 minutes. It is more consistent to get almost similar scores for both precision and recall, as shown in Figure 6.9. If we find the system defined boundary within $\pm/5$ minutes of the human judgement, then it is considered being accurate and we fixed that boundary as a segmentation point by the proposed sliding window algorithm (see Figure 6.8). Here, we calculated the precision in terms of true positives where

negative effect of over segmentation is ignored.

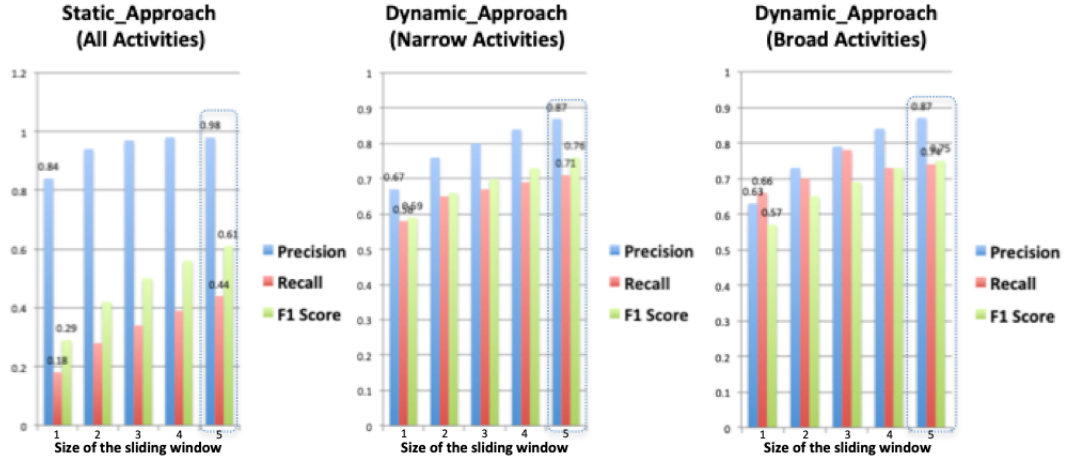


Figure 6.9: Identifying an optimal window size (i.e. $\pm[1$ to $5]$ minutes) in sliding window algorithm.

6.4.1 Exploring the Best Segmentation Algorithm

In the previous section, we focused on finding a few relevant documents for a given topic, especially in narrow-focused topics. In this experiment, we compare both the static and dynamic activity segmentation approaches for all 12 broad and 12 narrow focus queries to find the best lifelog segmentation algorithm. We employed sliding window approach as a tuning parameter in this experiment. We have used evaluation measures such as precision to identify the relevant documents accurately, recall ensuring the fraction of the successfully retrieved relevant documents, F1-score, and MCC score to observe overall performance of the proposed segmentation approaches.

- Static segmentation approach:** As explained in previous chapter, we used the sliding window approach to fine tune the proposed lifelog segmentation algorithm at indexing time. Referring to the results from Table 5.6 in the previous chapter, we get the score of precision (1), recall (0.55), F1-score (0.70), and MCC score (0.63). We found the best size of sliding window, i.e. ± 5 minutes in this approach. We describe the detailed discussion about the

results in a previous chapter and summarised in Table 6.6.

- **Dynamic segmentation approach:** The dynamic segmentation approach followed the methodology explained in the starting of this chapter and tested the segmentation algorithm using 12 broad and 12 narrow focused user query topics. From Table 6.5, we can observe that the precision and recall scores for both broad and narrow focused topics have increased. In this approach, we consider F1-score and MCC score as the measures of the overall accuracy of the proposed segmentation algorithm. We could achieve same accuracy performance in terms of F1-score i.e. 0.75 for both broad and narrow-focused topics; and MCC score i.e. 0.70 for broad and 0.65 for narrow-focused user query topics. However, we still need to pay extra attention to certain query topics such as cleaning (0.57), packing a suitcase (0.55), brainstorming (0.63), and ponytail (0.63), which are performing poorly, as can be observed in Table 6.5. The small number of relevant documents of such activities in the corpus could be the reason for this, which reduces the recall value and thus reflects on overall performance in terms of F1-score and MCC score of the segmentation algorithm.

Overall best segmentation approach: As we can observe from Table 6.6 that the overall accuracy in terms of average F1-score (0.75) and MCC score (0.70) for the dynamic approach outperforms the average F1-score (0.70) and MCC score (0.52) for the static approach by the improvement of 0.05 in terms of F1-score

Table 6.5: Accuracy scores for each activity in dynamic Segmentation approach.

Broad Topic User Queries					Narrow Topic User Queries				
Topics	Precision	Recall	F1-score	MCC score	Topics	Precision	Recall	F1-score	MCC score
working on laptop	0.94	0.82	0.85	0.91	walking to aeroplane	0.80	0.63	0.70	0.90
work meeting	0.74	0.70	0.70	0.87	walking at work	0.97	0.77	0.84	0.82
walking	0.91	0.96	0.92	0.85	brainstorming	0.56	0.72	0.63	0.79
relaxing at home	0.74	0.67	0.67	0.82	browsing items	0.96	0.85	0.88	0.78
socialising	0.79	0.79	0.76	0.78	packing	1	0.40	0.55	0.74
travelling in aeroplane	0.58	0.95	0.70	0.74	waiting for train	1	0.53	0.68	0.72
cooking	0.99	0.53	0.65	0.71	gardening	0.75	0.64	0.69	0.67
cleaning	1	0.44	0.57	0.65	writing on paper	1	0.95	0.97	0.59
driving	0.99	0.51	0.64	0.61	saturday morning coffee	0.58	0.73	0.64	0.50
retail shopping	0.95	0.77	0.84	0.55	fruit bowl	1	0.81	0.89	0.48
reading paper	0.94	0.89	0.91	0.52	eating apple	1	0.77	0.87	0.44
watching TV	0.82	0.83	0.79	0.47	ponytail	0.57	0.73	0.63	0.34
Average	0.87	0.74	0.75	0.70	Average	0.85	0.71	0.75	0.65

and 0.18 in terms of MCC score. Through experimentation, we compared both the segmentation approaches for all 24 user query topics. We explored the specific query based measure of improvement in terms of F1-score and MCC score for the dynamic approach against the conventional static approach. We identified the best lifelog segmentation approach, which justified our fifth research question i.e. RQ 5 i.e. Does a dynamic query-time based document segmentation approach outperforms a state-of-the-art indexing-time based lifelog segmentation approach for ad hoc lifelog retrieval.

Table 6.6: Performance of the static and the dynamic segmentation approaches.

Scores/Approach	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>MCC Score</i>
<i>Static</i>	1	0.55	0.70	0.52
<i>Dynamic (Narrow)</i>	0.85	0.71	0.75	0.65
<i>Dynamic (Broad)</i>	0.87	0.74	0.75	0.70

6.4.2 Discussion

In this section, we conclude that the dynamic segmentation approach outperforms static segmentation approach. We observed that mostly narrow-focused topics such as Saturday morning coffee, looking at fruit bowl, having a conversation with the person who has a ponytail, eating an apple and two broad-focused topics such as reading paper, and watching tv, have poor performance, as shown in Table 6.5. We can identify a number of reasons for this. Narrow-focused user queries topics are more specific to access user information which might be of short (one or a few minutes) duration. Thus, it makes the algorithm produce lower results in terms of recall to match the activity boundaries for such a small timestamp from huge corpus. It is our inference that the query topics should not be too specific so it could only retrieve one or few minutes of relevance judgements. We intend to explore this idea further that could provide useful insights to research community.

6.5 Chapter Summary

In this chapter, we introduced dynamic query-time based lifelog segmentation approaches. We improved the results by motivating the impact of manual annotations in the dynamic query-time based segmentation approach. We compared the both static and dynamic segmentation approaches to identify the better approach for activity-based lifelog segmentation. We tested both the segmentation algorithms using the reusable LSC2018 lifelog dataset. We appended manual annotations in feature vector and developed additional user query topics. We revised the existing sliding window approach [1] by employing different sizes of sliding window. Thus, we found that the dynamic activity based lifelog segmentation approach outperforms the conventional static activity-based lifelog segmentation approach. Hence, we justified our research hypothesis and resolved our fifth research question. In the next chapter, we will present the conclusion to this research work.

Chapter 7

Conclusion

7.1 Dissertation Overview

In this dissertation, we mainly focused on improving state-of-the-art lifelog segmentation approaches using publicly available lifelog datasets. According to the literature review presented in Chapter 2, we explored static indexing-time based segmentation approach to segment lifelog data. We also investigated the state-of-the-art lifelog segmentation approach based on early concept detector tools such as MPEG-7 [1] and developed a new approach employing high-quality visual features using Caffe framework [2] and Microsoft cognitive services API [3]. A significant outcome of this dissertation is a new dynamic query-time based segmentation algorithm to access and retrieve lifelog content as per query. A summary of the investigations, research findings, experimental results and proposed solutions in this dissertation are summarised as follows.

Chapter 1 provided an introduction of lifelogging and lifelog data, described motivations for this research, proposed hypothesis, listed research questions, major contributions of this dissertation and a list of publications. It also provided an outline of this dissertation.

Chapter 2 reviewed literature in lifelogging, use of digital technology in lifelogging, collection of lifelog data and techniques to process lifelog data. We reported the latest research work for both static indexing-time based and the dynamic query-time based segmentation approaches. Static indexing-time based segmentation indexes an individual's life activities with fixed duration. We found

the requirement of retrieving activities of an individual based on topic. We then outlined the pros and cons of both static indexing-time and dynamic query-time based segmentation algorithms. We concluded that detecting temporal event boundaries at indexing time remains a challenge. It motivated us to build a novel dynamic query-time based segmentation algorithm to improve the performance of segmentation algorithm.

Chapter 3 presented research methodology used for this research. We explained experimental methodology followed to retrieve meaningful information from lifelog data. We outlined generation and release of publicly available lifelog data and highlighted issues in selection of voluntary lifeloggers. This addressed second research question (RQ2) described in Chapter 1. We presented a user case study to observe considerations of privacy in lifelog datasets. This addressed first research question (RQ1) described in Chapter 1. We developed a hierarchy of different units of retrieval based on human memory explained in [17] to find the most appropriate unit for retrieving meaningful information using segmentation algorithm. This addressed the third research question (RQ3) described in Chapter 1.

Chapter 4 introduced lifelog data processing techniques. We analysed various methods related to data cleaning, data organisation and data extraction. We reported various data analytics techniques such as vector distance methods and automatic thresholding methods used in lifelog data analysis.

Chapter 5 presented the process of generating lifelog documents for segmentation task. This chapter focused on defining the baseline segmentation algorithm and introduced a preliminary experiment to segment visual lifelog data at indexing time. We proposed two novel approaches based on extracting high-quality visual features. We extracted visual features using Caffe framework and the Microsoft cognitive services API. We re-implemented baseline segmentation algorithm [1] to segment lifelog data into events. We used various data analytics technique such as the Euclidean vector distance method and

automatic thresholding techniques to highlight highly ranked event boundaries in lifelog data. Later in this chapter, we designed a segmentation algorithm based on activity retrieval unit and validated results using publicly available LSC2018 lifelog test collection for further experimentation. These experimentations addressed the fourth research question (RQ4) described in Chapter 1.

Chapter 6 focused on proposing the dynamic query-time based segmentation approach. We also showed the impact of using manual annotations with automatic annotations. Later in this chapter, we compared static indexing-time based and the dynamic query-time based segmentation algorithms where retrieval unit is an activity. Both segmentation approaches followed by methodology discussed in [1]. It validated our last research question RQ5 described in Chapter 1 and the hypothesis of this research. We summarise experiments conducted in Chapter 6:

In the first experiment, the authors used 15 regular user activities as query topics. NTCIR13-Lifelog evaluation workshop has introduced these activities under lifelog event segmentation (LEST) task. The authors suggested the need for developing new query topics to provide more specific information access to the user.

Second experiment considered importance of manual annotations in the dynamic query-time based segmentation. The authors introduced manual annotations based on a single level. The authors developed 24 new query topics which ranged from broad to narrow, focused type. This facilitated access to lifelog content as per information need. The authors found improved accuracy scores of precision using manual annotations.

In the third experiment, we utilised lifelog search challenge (LSC2018) dataset and manual annotations. We tested lifelog segmentation algorithm with precision, recall, F1-score and MCC metrics. We identified best size of a sliding window while fine-tuning the model. By comparing both static indexing-time based and the dynamic query-time based segmentation approaches, we found dynamic query-time based segmentation algorithm outperforms the static indexing-time based segmentation approach. It also justified the hypothesis of this dissertation.

7.2 Research Questions and Proposed Solutions

Hypotheses: *Dynamic (query-time) segmentation models for lifelog outperform conventional static (indexing-time) segmentation models for ad hoc search through lifelog datasets.*

From the hypotheses, we outlined the following research questions.

RQ1. What are the basic privacy concerns that researchers must address when gathering and analysing passively captured visual lifelog data and how researchers should use privacy-aware lifelog test collection to test different approaches to lifelog data segmentation and retrieval?

The lifelog data is highly sensitive and personal. We needed to consider the privacy of both lifelogger and bystanders as discussed in Chapter 3. We showed the need of preserving anonymity of lifelogger’s confidential data before releasing publicly. We presented a user case study to show privacy concerns of lifeloggers in sharing lifelog data. We defined a list of guidelines for researchers working or releasing lifelog data. Although it was a preliminary study, it allowed us to draw some initial useful conclusions for practitioners or researchers gathering lifelog datasets.

RQ2. How can we construct privacy-aware lifelog test collections to support comparative benchmarking for ad hoc retrieval tasks?

We addressed this research question in Chapter 3, where we investigated selection of lifeloggers and their associated privacy issues. We proposed principles of generating lifelog datasets such as continuity of data over a specified period that capture entirety of the user’s life experiences; the completeness of lifelog data can be attributed to the capture of four basic types of information: visual data, personal biometrics, human activity and information access; anonymity to remove any private user-identifiable

content from lifelog data and protectiveness to secure password protected storage of lifelog data. We also identified challenges in releasing privacy-aware lifelog datasets publicly. This research question resolved most of the challenges outlined in section 3.2.1 such as selection of voluntary lifeloggers; target audience and use-case; wearable devices to collect lifelog data and storage of lifelog data.

RQ3. What are various units of retrieval documents for ad hoc search through multimodal lifelog data?

We investigated this research question in Chapter 3 while presenting lifelog data processing techniques. We proposed a hierarchy of the retrieval units based on the five R's of memory access introduced by Sellen and Whittaker in 2010 [17]. We identified 'activity' as the most appropriate basic unit of retrieval for this research. Through experimentation in chapter 5, we observed that segmentation algorithm based on activity retrieval as unit performed better. We link these results to the following research questions and hypothesis of this dissertation.

RQ4. What is the state-of-the-art indexing time segmentation approach?

For this research question, we initially explored process of generating lifelog documents based on implementing findings from previous research questions in Chapter 5. We addressed up-to-date literature review to explore existing approaches to segment lifelog data. We defined baseline segmentation algorithm based on a literature review. This research question was investigated by presenting a preliminary user experiment described in Chapter 5. Two novel segmentation approaches based on extracting visual features from Caffe framework and Microsoft cognitive services API) were proposed and compared with baseline segmentation [1] approach. This improved overall system performance by 0.36. We fixed this approach as our baseline approach for building segmentation algorithms. We evaluated and

justified our proposed segmentation algorithm using LSC2018 lifelog test collection later in this chapter.

RQ5. Does a dynamic query-time based document segmentation approach outperforms a state-of-the-art indexing-time based lifelog segmentation approach for ad hoc lifelog retrieval?

This research question directly links to our hypothesis and it is investigated in Chapter 6. We re-implemented experimental methodology to test dynamic query-time based segmentation algorithms discussed in chapter 5. We presented a preliminary experiment where we designed a dynamic query-time based segmentation algorithm using query topics introduced in LEST task. These topics were further revised by using manual annotations. It was refined by building new query topics specific to irregular human activities. We compared both the dynamic query-time based and static indexing-time based approaches and found that dynamic segmentation approach outperformed static segmentation approach. This answered the fifth research question. We could return to the hypothesis of this dissertation and prove it.

7.3 Contributions of this Dissertation

Following are the contributions we have justified and explained:

- [1] Developed the process of releasing privacy-aware lifelog test collections for continuous comparative evaluation in retrieval tasks (RQ1 and RQ2).
- [2] Developed an ontology for the basic units of retrieval based on 5 R's of memory access [17] in lifelogging and define optimal size of lifelog documents (RQ3).
- [3] Developed indexing-time based segmentation approaches by using high-quality visual features and data analytic techniques to segment lifelog data accurately (RQ4).

- [4] Investigated and developed dynamic query-time based segmentation algorithms to provide the flexible segmentation mechanism as per information needs (RQ5). In addition, we investigated the most appropriate size of window in a sliding window approach to improve the accuracy of the segmentation algorithm.

7.4 Limitations of this Work

- [1] **Small dataset and limit of cohort diversity:** We found many limitations from our user study on preserving anonymity of lifelog data in Chapter 3. Although the cohort is preliminary, we draw conclusions when we examine inter-cohort differences in terms of profession. Some lifeloggers might have filtered their lifelog data before starting the experiment is another limitation. This was not deemed as a major issue.

We suggest that we need to utilise larger and more diverse cohort so we can validate initial findings and engage in a more in-depth analysis. We have only considered visual lifelog image data for this experiment. It should be noted that there might be other privacy concerns raised by other data sources such as location, heart rate, URLs visited while addressing lifelog data anonymisation. Therefore, we suggest building an automated anonymisation system. This anonymisation tool would include a list of data governance guidelines and data privacy ethics. It would also include the consent form for third party organisations access, usage of automatic blurring tools to blur identifiable content and a survey form for users to know personal experience while collecting lifelog data.

- [2] **Quality of automatic annotation:** It is clear in chapter 5 and 6 that quality of automatic annotations varies with the visual concept detector tool being used. There would always be a difference between the low-quality e.g. MPEG-7 features and high-quality e.g. Caffe visual concepts and image

categories. As discussed in Chapter 5, the low-quality automatic annotations are machine readable and thus much faster to implement. High-quality automatic annotations are semantically enriched and humanly understandable. Therefore, we reported that there would always be a semantic gap between the low and high quality automatic annotations.

- [3] **Challenge of comparing the static and the dynamic segmentation approaches:** It was a challenge to compare static and dynamic segmentation algorithms for each specific query topic. As we know the nature of static segmentation approach where we define user activities at indexing time and cannot be changed later. In dynamic segmentation approach, after user activities are defined at indexing time, we can retrieve activities at retrieval time as per information needs. We solved this problem by calculating accuracy score for each specific query topic and average the accuracy scores to get an overall performance of proposed dynamic query-time based segmentation algorithm. We were then in the position to compare these results with the results stated for static indexing-time based segmentation approach in chapter 5.

7.5 Suggestions for the Future Research

There are several potential directions for the future research and further development of this research. We outline some interesting extensions and prospects as follows.

- [1] **High volumes of multimodal lifelog data:** It is our belief that we should gather data from multiple sensors to capture entirety of an individual life experience. Digital record of such a large volume of personal content will enable retrieval of useful information over the years. Therefore, we can suggest the research community to gather a large volume of lifelog data to retrieve any kind of user life activity. For example, if we need lifelog data to analyse personal dietary routine habits of an individual or a group of individuals, we

must gather a food log over a period of one month. The authors believe that one month is sufficient duration of analysing someone's food habits.

- [2] **Willingness of the participants for the experimentation:** We believe that participants should voluntary log their own personal digital record. They must be familiar with gathering lifelog data and understand the presence of other people in their lifelogs. However, finding such people willing to devote months or years or decades or even their entire life will always remain a challenge in lifelogging.

- [3] **Building automatic secure storage system for lifelog data:** Increased advancement and availability of cost effective storage devices in market has solved the problem of storing large volume of lifelog data for a longer period. We need to build a secure and fast storage system to store lifelog data. The author also examined the need to build fast wireless systems which can be installed on lifelogging devices to transmit data to these cloud based storage. We believe that it is inefficient and time consuming for the lifelogger to send data to the cloud server. The dissertation writer investigated the need of building such automatic secure storage system that automatically syncs lifelog sensors on specified time interval i.e. every 2 hours via wireless connection and secure all sensor data by allowing it with specific user password.

- [4] **Building an information retrieval engine using proposed activity based dynamic (query-time) segmentation approach:** Building an ad hoc search engine based on the proposed activity based dynamic segmentation algorithm could become major future endeavour. Considering a specific case scenario where an individual want to analyse behavioural change over time i.e. over the period of 20 to 30 years. This could be done for a variety of use-cases such as keeping track for personal well-being or maintaining progress report of an athlete for performance analysis.

7.6 Concluding Remarks

It has become easy to record lifelog data digitally in past decade. It may be worthwhile to gather as much as information as suggested by O'Hara et al. [98] as “every piece of information may be valuable”. We have found that secure storage and data governance should be considered as the highest priority in lifelogging. However, one caveat to note here is that even if we perform experiment anonymously, we do not know the risk of identification of participant in future. So authors suggested erring on the side of caution and to assume that we know the identity of a lifelogger in a released dataset or collection, regardless of the steps taken to ensure anonymity. We defined concept of generating lifelog documents and defined most appropriate unit of lifelogging data retrieval. We explored existing static indexing-time based segmentation approaches developed until now and we developed a new algorithm based on our baseline segmentation approach. We used this algorithm for building a dynamic query-time based segmentation algorithm. We believe that the proposed segmentation algorithm would provide a good building block to develop ad hoc search engines for lifelog datasets and future researcher can get valuable insights and guidance for their research.

Bibliography

- [1] A. R. Doherty and A. F. Smeaton. Automatically segmenting lifelog data into events. *9th International Workshop on Image Analysis for Multimedia Interactive Services*, 30 June 2008.
- [2] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.
- [3] Hao Fang, Saurabh Gupta, Forrest Iandola, Rupesh Srivastava, Li Deng, Piotr Dollar, Jianfeng Gao, Xiaodong He, Margaret Mitchell, John Platt, Lawrence Zitnick, and Geoffrey Zweig. From captions to visual concepts and back. *IEEE Institute of Electrical and Electronics Engineers*, June 2015.
- [4] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. Lifelogging: Personal big data. *Foundations and Trends in Information Retrieval*, 8(1):1–125, 2014.
- [5] Martin Dodge and Kitchin Rob. Outlines of a world coming into existence’: Pervasive computing and the ethics of forgetting. *Environment and Planning B: Planning and Design*, 34:431–445, June 2007.
- [6] Vannevar Bush. *The Atlantic Monthly*, (1):101–108, Jul.
- [7] Gordon Bell and Jim Gemmell. A digital life. *Scientific American*, 296:58–65, February 2007.
- [8] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Rashmi Gupta, Rami Albatal, and Duc Tien Dang Nguyen. Overview of ntcir-13 lifelog-2 task. *The Thirteenth NTCIR conference (NTCIR-13)*, dec 2017.

- [9] Zaher Hinbarji, Rami Albatal, Noel O'Connor, and Cathal Gurrin. Loggerman, a comprehensive logging and visualization tool to capture computer usage. In *MultiMedia Modeling*, pages 342–347, Cham, 2016. Springer International Publishing.
- [10] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and title= Albatal, Rami.
- [11] Rashmi Gupta. Considering documents in lifelog information retrieval. *ICMR'18. In: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, page 497–500, 2018.
- [12] Aiden R. Doherty, Alan F. Smeaton, Keansub Lee, and Danial P.W. Ellis. Multimodal segmentation of lifelog data. *Large Scale Semantic Access to Content (Text, Image, Video, and Sound)*, pages 21–38, june 2007.
- [13] Yi Chen, Gareth J.F. Jones, and Debasis Ganguly. Segmenting and summarizing general events in a long-term lifelog. In: *The 2nd Workshop Information Access for Personal Media Archives (IAPMA) at ECIR 2011*, April 2011.
- [14] Zhen Li, Zhiqiang Wei, Wenyan Jia, and Mingui Sun. Daily life event segmentation for lifestyle evaluation based on multi-sensor data recorded by a wearable device. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2858–2861, 2013.
- [15] Marc Bolanos, Ricard Mestre, Estefania Talavera, Xavier G. Nieto, and Petia Radeva. Visual summary of egocentric photostreams by representative keyframes. *IEEE First International Workshop on Wearable and Ego-vision Systems for Augmented Experience (WEsAX)*, 29 June - 3 July, 2015, Turin, Italy.
- [16] Daragh Byrne, Barry Lavelle, Aiden R. Doherty, Gareth J.F. Jones, and Alan F. Smeaton. Using bluetooth and gps metadata to measure event

similarity in sensecam images. *5th International Conference on Intelligent Multimedia and Ambient Intelligence*, July 2007.

- [17] Abigail Sellen and Steve Whittaker. Beyond total capture: A constructive critique of lifelogging. *Communications of the ACM*, 53:70–77, may 2010.
- [18] Duc-Tien Dang-Nguyen, Luca Piras, Michael Riegler, Liting Zhou, Mathias Lux, and Cathal Gurrin. Overview of imagecleflifelog 2018: Daily living understanding and lifelog moment retrieval. In *CLEF*, 2018.
- [19] D.-T. Dang Nguyen, K. Schoeffmann, and W. Hurst. Lsc2018 panel-challenges of lifelog search and access. *ACM Workshop on The Lifelog Search Challenge*, 11-14 June 2018.
- [20] Marc Bolaños, Mariella Dimiccoli, and Petia Radeva. Towards storytelling from visual lifelogging: An overview. *CoRR*, abs/1507.06120, 2015.
- [21] Sudheendra Hangal and Monica S. Lam. Sentiment analysis on personal email archives, 2011.
- [22] Matthew L. Lee and Anind K. Dey. Using lifelogging to support recollection for people with episodic memory impairment and their caregivers. In *Proceedings of the 2Nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments*, HealthNet '08, pages 14:1–14:3, New York, NY, USA, 2008. ACM.
- [23] Matthew L. Lee and Anind K. Dey. Lifelogging memory appliance for people with episodic memory impairment. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, UbiComp '08, pages 44–53, New York, NY, USA, 2008. ACM.
- [24] Marc Bolaños, Maite Garolera, and Petia Radeva. Active labeling application applied to food-related object recognition. In *Proceedings of the 5th*

International Workshop on Multimedia for Cooking & Eating Activities, CEA '13, pages 45–50, New York, NY, USA, 2013. ACM.

- [25] Edison Thomaz. Practical food journaling. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, UbiComp '13 Adjunct, pages 355–360, New York, NY, USA, 2013. ACM.
- [26] Hyungik Oh, Jonathan Nguyen, Soundarya Soundararajan, and Ramesh Jain. Multimodal food journaling. In *Proceedings of the 3rd International Workshop on Multimedia for Personal Health and Health Care*, HealthMedia'18, pages 39–47, New York, NY, USA, 2018. ACM.
- [27] Floriano Zini, Martin Reinstadler, and Francesco Ricci. Increasing quality of life awareness with life-logging. In Kostas Giokas, Laszlo Bokor, and Frank Hopfgartner, editors, *eHealth 360*, pages 282–293, Cham, 2017. Springer International Publishing.
- [28] Smith Ashley, O'Hara Kieron, and Lewis Paul. Visualising the past: Annotating a life with linked open data. June 2011. Event Dates: June 14th-17th 2011.
- [29] J. Gemmell, G. Bell, and R. Lueder. MyLifeBits: a personal database for everything. *Communications of the ACM*, 49(1):88–95, 2006.
- [30] Han-Saem Park and Sung-Bae Cho. Building mobile social network with semantic relation using bayesian network-based life-log mining. pages 401–406, 08 2010.
- [31] Yuri Yoshihara, Dalai Tang, and Naoyuki Kubota. Life log visualization system based on informationally structured space for supporting elderly people. *2013 Second International Conference on Robot, Vision and Signal Processing*, pages 78–83, 2013.

- [32] Yvette E. Gelogo and Haeng-Kon Kim. Integration of wearable monitoring device and android smartphone apps for u-healthcare monitoring system. 2015.
- [33] Frank Nack. You must remember this. *IEEE MultiMedia*, 12(1):4–7, 2005.
- [34] Steve Mann. Wearable computing : A first step toward personal imaging. *Personal and Ubiquitous Computing*, 30, 1997.
- [35] Shuaib Karim, Amin Andjomshoaa, and A. Min Tjoa. Exploiting sensecam for helping the blind in business negotiations. In Klaus Miesenberger, Joachim Klaus, Wolfgang L. Zagler, and Arthur I. Karshmer, editors, *Computers Helping People with Special Needs*, pages 1147–1154, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
- [36] Daragh Byrne and Gareth J. F. Jones. Creating stories for reflection from multimodal lifelogcontent: An initial investigation. 2009.
- [37] Ali Mair, Marie Poirier, and Martin A Conway. Memory for staged events: Supporting older and younger adults’ memory with sensecam. *Quarterly Journal of Experimental Psychology*, 72(4):717–728, 2019. PMID: 29504463.
- [38] Soumyadeb Chowdhury, Md. Sadek Ferdous, and Joemon M. Jose. A user-study examining visualization of lifelogs. *2016 14th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pages 1–6, 2016.
- [39] Louise N Signal, James Stanley, Moira Beatrice Smith, Michael Barr, Tim J. Chambers, Jiang Zhou, Aaron Duane, Cathal Gurrin, Alan F. Smeaton, Christina McKerchar, Amber L Pearson, Janet A Hoek, Gabrielle Jenkin, and Cliona Ni Mhurchu. Children’s everyday exposure to food marketing: an objective analysis using wearable cameras. In *The international journal of behavioral nutrition and physical activity*, 2017.

- [40] Chenyou Fan, Zehua Zhang, and David J. Crandall. Deepdiary: Lifelogging image captioning and summarization. *Journal of Visual Communication and Image Representation*, 55:40 – 55, 2018.
- [41] Shiqi Jiang, Zhenjiang Li, Pengfei Zhou, and Mo Li. Memento: An emotion-driven lifelogging system with wearables. *ACM Transactions on Sensor Networks*, 15:1–23, 01 2019.
- [42] Ana Garcia del Molino, Cheston Tan, Joo-Hwee Lim, and Ah-Hwee Tan. Summarization of egocentric videos: A comprehensive survey. *IEEE Transactions on Human-Machine Systems*, 47:65–76, 2017.
- [43] Francesco Ragusa, Antonino Furnari, Sebastiano Battiato, Giovanni Signorello, and Giovanni Maria Farinella. Egocentric visitors localization in cultural sites. *J. Comput. Cult. Herit.*, 12(2):11:1–11:19, April 2019.
- [44] Keith Brazendale, Michael W. Beets, Robert G Weaver, Michelle W. Perry, Emily B. Tyler, Ethan T Hunt, Lindsay K. Decker, and J. D. Chaput. Comparing measures of free-living sleep in school-aged children. *Sleep medicine*, 2019.
- [45] J. Dunn, R. Runge, and M. Snyder. Wearables and the medical revolution. *Personalized Medicine*, 15(5):429–448, 2018. cited By 7.
- [46] M. de Zambotti, L. Rosas, I.M. Colrain, and F.C. Baker. The sleep of the ring: Comparison of the Ōura sleep tracker against polysomnography. *Behavioral Sleep Medicine*, 17(2):124–136, 2019. cited By 6.
- [47] Zaher Hinbarji, Rami Albatal, Noel E. O’Connor, and Cathal Gurrin. Loggerman, a comprehensive logging and visualisation tool to capture computer usage. In: *22st International Conference on MultiMedia Modelling*, pages 342–347, 2016.

- [48] Grace Therrell. More product, more process: metadata in digital image collections. *Digital Library Perspectives*, 35(1):2–14, 2019.
- [49] Anind K. Dey. Understanding and using context. *Personal Ubiquitous Comput.*, 5(1):4–7, January 2001.
- [50] A. Cho, H. Lee, H. Hwang, Y. Jo, and M. Whang. Embodied emotion recognition system. *Lecture Notes in Electrical Engineering*, 474:1371–1376, 2018. cited By 0.
- [51] P. Herruzo, M. Bolaños, and P. Radeva. Can a cnn recognize catalan diet? volume 1773, 2016. cited By 0.
- [52] Ling Bao and Stephen S. Intille. Activity recognition from user-annotated acceleration data. In Alois Ferscha and Friedemann Mattern, editors, *Pervasive Computing*, pages 1–17, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.
- [53] Zhen-Yu He and Lian-Wen Jin. Activity recognition from acceleration data using ar model representation and svm. volume 4, pages 2245 – 2250, 08 2008.
- [54] Tzu-Ping Kao, Che-Wei Lin, and Jeen-Shing Wang. Development of a portable activity detector for daily activity recognition. pages 115 – 120, 08 2009.
- [55] Martin Berchtold, Matthias Budde, Dawud Gordon, Hedda Rahel Schmidtke, and Michael Beigl. Actiserv: Activity recognition service for mobile phones. *International Symposium on Wearable Computers (ISWC) 2010*, pages 1–8, 2010.
- [56] Myong-Woo Lee, Adil Mehmood Khan, and Tae-Seong Kim. A single tri-axial accelerometer-based real-time personal life log system capable of human activity recognition and exercise information generation. *Personal and Ubiquitous Computing*, 15(8):887–898, Dec 2011.

- [57] Zhixian Yan, Vigneshwaran Subbaraju, Dipanjan Chakraborty, Archan Misra, and Karl Aberer. Energy-efficient continuous activity recognition on mobile phones: An activity-adaptive approach. In *Proceedings of the 2012 16th Annual International Symposium on Wearable Computers (ISWC)*, ISWC '12, pages 17–24, Washington, DC, USA, 2012. IEEE Computer Society.
- [58] E. Spyrou, G. Tolias, P. Mylonas, and Y. Avrithis. Concept detection and keyframe extraction using a visual thesaurus. *Multimedia Tools and Applications*, 41(3):337–373, February 2009.
- [59] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.
- [60] Utz Westermann and Ramesh Jain. Events in multimedia electronic chronicles (e-chronicles). *Int. J. Semantic Web Inf. Syst.*, 2:1–23, 2006.
- [61] Hyowon Lee, Alan F. Smeaton, Noel E. O'Connor, Gareth J. F. Jones, Michael Blighe, Daragh Byrne, Aiden R. Doherty, and Cathal Gurrin. Constructing a sensecam visual diary as a media process. *Multimedia Systems*, 14:341–349, 2008.
- [62] Liadh Kelly and Gareth Jones. Biometric response as a source of query independent scoring in lifelog retrieval. 03 2010.
- [63] Stephen Robertson and Hugo Zaragoza. The probabilistic relevance framework: Bm25 and beyond. *Found. Trends Inf. Retr.*, 3(4):333–389, April 2009.
- [64] Snehal Chennuru, Peng-Wen Chen, Jiang Zhu, and Ying Zhang. Mobile lifelogger – recording, indexing, and understanding a mobile user’s life. 01 2012.
- [65] Lijuan Marissa Zhou, Cathal Gurrin, and Zhengwei Qiu. Zhiwo: Activity tagging and recognition system for personal lifelogs. In *Proceedings of the*

3rd ACM Conference on International Conference on Multimedia Retrieval, ICMR '13, pages 321–322, New York, NY, USA, 2013. ACM.

- [66] Seyed Ali Bahrainian and Fabio Crestani. Cued retrieval of personal memories of social interactions. In *Proceedings of the First Workshop on Lifelogging Tools and Applications*, LTA '16, pages 3–12, New York, NY, USA, 2016. ACM.
- [67] Bogdan Ionescu, Henning Müller, Mauricio Villegas, Helbert Arenas, Giulia Boato, Duc-Tien Dang-Nguyen, Yashin Dicente Cid, Carsten Eickhoff, Alba G. Seco de Herrera, Cathal Gurrin, Bayzidul Islam, Vassili Kovalev, Vitali Liauchuk, Josiane Mothe, Luca Piras, Michael Riegler, and Immanuel Schwall. Overview of imageclef 2017: Information extraction from images. In Gareth J.F. Jones, Séamus Lawless, Julio Gonzalo, Liadh Kelly, Lorraine Goeuriot, Thomas Mandl, Linda Cappellato, and Nicola Ferro, editors, *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pages 315–337, Cham, 2017. Springer International Publishing.
- [68] Bogdan Ionescu, Henning Müller, Mauricio Villegas, Alba García Seco de Herrera, Carsten Eickhoff, Vincent Andrearczyk, Yashin Dicente Cid, Vitali Liauchuk, Vassili Kovalev, Sadid A. Hasan, Yuan Ling, Oladimeji Farri, Joey Liu, Matthew Lungren, Duc-Tien Dang-Nguyen, Luca Piras, Michael Riegler, Liting Zhou, Mathias Lux, and Cathal Gurrin. Overview of imageclef 2018: Challenges, datasets and evaluation. In Patrice Bellot, Chiraz Trabelsi, Josiane Mothe, Fionn Murtagh, Jian Yun Nie, Laure Soulier, Eric SanJuan, Linda Cappellato, and Nicola Ferro, editors, *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pages 309–334, Cham, 2018. Springer International Publishing.
- [69] Bogdan Ionescu, Henning Müller, Renaud Péteri, Duc-Tien Dang-Nguyen, Luca Piras, Michael Riegler, Minh-Triet Tran, Mathias Lux, Cathal Gurrin, Yashin Dicente Cid, Vitali Liauchuk, Vassili Kovalev, Asma Ben Abacha,

Sadid A. Hasan, Vivek Datla, Joey Liu, Dina Demner-Fushman, Obioma Pelka, Christoph M. Friedrich, Jon Chamberlain, Adrian Clark, Alba García Seco de Herrera, Narciso Garcia, Ergina Kavallieratou, Carlos Roberto del Blanco, Carlos Cuevas Rodríguez, Nikos Vasillopoulos, and Konstantinos Karampidis. Imageclef 2019: Multimedia retrieval in lifelogging, medical, nature, and security applications. In Leif Azzopardi, Benno Stein, Norbert Fuhr, Philipp Mayr, Claudia Hauff, and Djoerd Hiemstra, editors, *Advances in Information Retrieval*, pages 301–308, Cham, 2019. Springer International Publishing.

- [70] Liting Zhou, Zaher Hinbarji, Duc-Tien Dang-Nguyen, and Cathal Gurrin. Lifer: An interactive lifelog retrieval system. In *Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge*, LSC '18, pages 9–14, New York, NY, USA, 2018. ACM.
- [71] Bernd Münzer, Andreas Leibetseder, Sabrina Kletz, Manfred Jürgen Primus, and Klaus Schoeffmann. lifexplore at the lifelog search challenge 2018. In *Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge*, LSC '18, pages 3–8, New York, NY, USA, 2018. ACM.
- [72] Jakub Lokoč, Tomáš Souček, and Gregor Kovalčík. Using an interactive video retrieval tool for lifelog data. In *Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge*, LSC '18, pages 15–19, New York, NY, USA, 2018. ACM.
- [73] Matthew J. Lavin. Analyzing documents with tf-idf,. *The Programming Historian* 8 (2019), June 2019.
- [74] Tie-Yan Liu. Learning to rank for information retrieval. *Found. Trends Inf. Retr.*, 3(3):225–331, March 2009.
- [75] Donna Harman. Overview of the first trec conference. In *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and*

Development in Information Retrieval, SIGIR '93, pages 36–47, New York, NY, USA, 1993. ACM.

- [76] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Van-Tu Ninh, Tu-Khiem Le, Rami Albatal, Duc Tien Dang Nguyen, and Graham Healy. Overview of the ntcir-14 lifelog-3 task. *The 14th NTCIR conference (NTCIR-14)*, June 2019.
- [77] Van-Tu Ninh, Tu-Khiem Le, Liting Zhou, Graham Healy, Kaushik, Venkataraman, Minh-Triet Tran, Duc-Tien Dang-Nguyen, Sinéad Smyth, and Cathal Gurrin. A baseline interactive retrieval engine for the ntcir-14 lifelog-3 semantic access task.
- [78] Cathal Gurrin, Klaus Schoeffmann, Hideo Joho, Andreas Leibetseder, Liting Zhou, Aaron Duane, Duc-Tien Dang-Nguyen, Michael Riegler, Luca Piras, Minh-Triet Tran, Jakub Lokoč, and Wolfgang Hürst. Comparing Approaches to Interactive Lifelog Search at the Lifelog Search Challenge (LSC2018). *ITE Transactions on Media Technology and Applications*, 7(2):46–59, 2019.
- [79] Duc-Tien Dang-Nguyen, Liting Zhou, Rashmi Gupta, Michael Riegler, and Cathal Gurrin. Building a disclosed lifelog dataset: Challenges, principles and processes. In *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*, CBMI '17, pages 22:1–22:6, New York, NY, USA, 2017. ACM.
- [80] Rashmi Gupta and Cathal Gurrin. Approaches for event segmentation of visual lifelog data. In *Schoeffmann K. et al. (eds) MultiMedia Modeling. MMM 2018. Lecture Notes in Computer Science, vol 10704. Springer, Cham*, 2018.
- [81] Nicola Stokes, Joe Carthy, and Alan F. Smeaton. Select: A lexical cohesion based news story segmentation system. *AI Commun.*, 17(1):3–12, January 2004.

- [82] D. Chakrabarti and K. Punera. Event summarization using tweets. In *Proc. 6th AAAI Int. Conf. on Weblogs and Social Media*, 2011.
- [83] James Allan. *Topic Detection and Tracking: Event-based Information Organization*. Springer Publishing Company, Incorporated, 2012.
- [84] Stanley Boykin and Andrew Merlino. Machine learning of event segmentation for news on demand. *Commun. ACM*, 43(2):35–41, February 2000.
- [85] Svebor Karaman, Jenny Benois-Pineau, Vladislavs Dovgalecs, Rémi Mégret, Julien Pinquier, Régine André-Obrecht, Yann Gaëstel, and Jean-François Dartigues. Hierarchical hidden markov model in detecting activities of daily living in wearable videos for studies of dementia. *Multimedia Tools and Applications*, 69(3):743–771, Apr 2014.
- [86] Ekaterina H. Spriggs, Fernando De la Torre, and Martial Hebert. Temporal segmentation and activity classification from first-person sensing. *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 17–24, 2009.
- [87] Marc Bolaños, Maite Garolera, and Petia Radeva. Video segmentation of life-logging videos. In Francisco José Perales and José Santos-Victor, editors, *Articulated Motion and Deformable Objects*, pages 1–9, Cham, 2014. Springer International Publishing.
- [88] Yair Poleg, Chetan Arora, and Shmuel Peleg. Temporal segmentation of egocentric videos. In *CVPR*, 2014.
- [89] Antonino Furnari, Sebastiano Battiato, and Giovanni Maria Farinella. Personal-location-based temporal segmentation of egocentric videos for lifelogging applications. *J. Visual Communication and Image Representation*, 52:1–12, 2018.

- [90] Raja Ghazi, Olfa Fraj, and Meriem Jaïdane. Visually-based audio texture segmentation for audio scene analysis. *2007 15th European Signal Processing Conference*, pages 1531–1535, 2007.
- [91] S.E. Kucukbay and Mustafa Sert. Audio event detection using adaptive feature extraction scheme. In *MMEDIA 2015, The Seventh International Conferences on Advances in Multimedia*, pages 44–49, April 2015.
- [92] D. A. Sadlier, Niall O’Connor, Seán Marlow, and Noel Murphy. A combined audio-visual contribution to event detection in field sports broadcast video. case study: Gaelic football. *Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology (IEEE Cat. No.03EX795)*, pages 552–555, 2003.
- [93] Qin Jin, Kornel Laskowski, Tanja Schultz, and Alex Waibel. Speaker segmentation and clustering in meetings. 01 2004.
- [94] Marti A. Hearst and Christian Plaunt. Subtopic structuring for full-length document access. In *SIGIR*, 1993.
- [95] Shûhei Yamamoto, Takuya Nishimura, Yoshiaki Takimoto, Takafumi Inoue, and Hiroyuki Toda. Pbg at the ntcir-13 lifelog-2 lat , lsat , and lest tasks. 2017.
- [96] Ana Garcia del Molino, Joo-Hwee Lim, and Ah-Hwee Tan. Predicting visual context for unsupervised event segmentation in continuous photo-streams. In *Proceedings of the 26th ACM International Conference on Multimedia*, MM ’18, pages 10–17, New York, NY, USA, 2018. ACM.
- [97] Melanie Swan. The quantified self: Fundamental disruption in big data science and biological discovery. *Big data*, 1 2:85–99, 2013.

- [98] Kieron O’Hara, Mischa M. Tuffield, and Nigel Shadbolt. Lifelogging: Privacy and empowerment with memories for life. *Identity in the Information Society*, 1(1):155–172, Dec 2008.
- [99] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatal. Nteir lifelog: The first test collection for lifelog research. In *SIGIR*, 2016.
- [100] L. Trom and J. Cronje. Analysis of data governance implications on big data. *Lecture Notes in Networks and Systems*, 69:645–654, 2020. cited By 0.
- [101] A. Zaheer, B. McEvily, and V. Perrone. Does trust matter? exploring the effects of interorganizational and interpersonal trust on performance. *Organization Science*, 9(2):141–159, 1998. cited By 2032.
- [102] Paul Kelly, Simon J. Marshall, Hannah Badland, Jacqueline Kerr, Melody Oliver, Aiden R. Doherty, and Charlie Foster. An ethical framework for automated, wearable cameras in health behavior research. *American Journal of Preventive Medicine*, 44(3):314 – 319, 2013.
- [103] Rashmi Gupta, Martin Crane, and Cathal Gurrin. Considerations on privacy in the era of digitally logged lives. *Online Information Review*, December 2020.
- [104] Soumyadeb Chowdhury, Md Sadek ferdous, and Joemon M Jose. Exploring lifelog sharing and privacy. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 553–558, 2016.
- [105] Soumyadeb Chowdhury, Md Sadek ferdous, and Joemon M Jose. Understanding lifelog sharing preferences of lifeloggers. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*, pages 649–651, 2016.

- [106] Joseph Phelps, Glen Nowak, and Elizabeth Ferrell. Privacy concerns and consumer willingness to provide personal information. *Journal of Public Policy & Marketing, Privacy and Ethical Issues in Database/Interactive Marketing and Public Policy*, 19(1):27–41, 2000.
- [107] Antonio Leucio Cutillo, Refik Molva, and Thorsten Strufe. Safebook: A privacy-preserving online social network leveraging on real-life trust. In *IEEE Communications Magazine, December*, pages 94–101, 2009.
- [108] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatal. 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*, pages 705–708, 2016.
- [109] Roberto Hoyle, Robert Templeman, Dennis Anthony, David Crandall, and Apu Kapadia. Sensitive lifelogs: A privacy analysis of photos from wearable cameras. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1645–1648, 2015.
- [110] Robert Byrne and Dermot Diamond. Chemo/bio-sensor networks. *Nature materials*, 5(6):421—424, June 2006.
- [111] Aaron Duane, Rashmi Gupta, Liting Zhou, and Cathal Gurrin. Visual insights from personal lifelogs: Insight at the ntcir-12 lifelog lit task. In *NTCIR*, 2016.
- [112] Anita L. Allen. Dredging-up the past: Lifelogging, memory and surveillance. *Scholarship at Penn Law*, 75:47–74, 12 2008.
- [113] Maarten den Braber. 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*, October 2016.

- [114] Li Zhen, Wei Zhiqiang, Jia Wenyan, and Sun Mingui. 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.
- [115] Daniel Kahneman, Alan B. Krueger, David A. Schkade, Norbert Schwarz, and Arthur A. Stone. A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306:1776–1780, 2004.
- [116] Rami Albatal, Cathal Gurrin, Jiang Zhou, Yang Yang, Denise Carthy, and Na Li. Senseseer mobile-cloud-based lifelogging framework. In *ISTAS*, pages 144–146. IEEE, 2013.
- [117] Minh-Son Dao, Duc-Tien Dang-Nguyen, Michael Riegler, and Cathal Gurrin. Smart lifelogging: Recognizing human activities using phasor. In *ICPRAM*, 2017.
- [118] Louise Signal, Moira Smith, Michelle Stanley, James Chambers, Tim Barr, Mhurchu, Jiang Zhou, Aaron Duane, Gabrielle Jenkin, Amber L Pearson, Cathal Gurrin, Alan F. Smeaton, Janet Hoek, and Cliona Ni Mhurchu. Semantic indexing of wearable camera images: Kids’cam concepts. In *American journal of preventive medicine, April 2017*, April, 2017.
- [119] Yael Onn, Yaniv Druchman, Rom Timor, Arz Maroun, Yossi Nachmani, Saar Sichlai, Maor Fishman, Michael Geva, Ariel Zyssman, Inbal Lev, Tamar Maron, Yaniv Simsolo, Adi Fuches, and Shai Packer. Privacy in the digital environment. 2005.
- [120] Alan F. Westin. Privacy and freedom. In *Washington and Lee Law Review*, 25(1), 166.
- [121] Darhl M. Pedersen. Dimensions of privacy. In *Perceptual and Motor Skills*, 48(3c), pages 1291–1297, 1979.

- [122] Mark S Ackerman and Scott D. Mainwaring. Privacy issues and human-computer interaction. pages 19–26, 2005.
- [123] Kieron O’Hara, Mischa M. Tuffield, and Nigel Shadbolt. Lifelogging: Privacy and empowerment with memories for life. In *Identity in the Information Society, 1 (155)*, pages 155–172, 2008.
- [124] Cathal Gurrin, Rami Albatal, Hideo Joho, and Kaori Ishii. A privacy by design approach to lifelogging. In *In: O Hara, K. and Nguyen, C. and Haynes, P., (eds.) Digital Enlightenment Yearbook, 2014*, pages 49–73.
- [125] Tengqi Ye, Brian Moynagh, Rami Albatal, and Cathal Gurrin. Negative faceblurring: A privacy-by-design approach to visual lifelogging with google glass. In *Proceedings of the 2014 ACM International Conference on Information and Knowledge Management*, pages 2036–2038, 2014.
- [126] Mohammed Korayem, Robert Templeman, Dennis Chen, David Crandall, and Apu Kapadia. Enhancing lifelogging privacy by detecting screens. In *Conference on Human Factors in Computing Systems*, pages 4309–4314, 2016.
- [127] Martin Dodge and Rob Kitchin. ’outlines of a world coming into existence’: Pervasive computing and the ethics of forgetting. *Environment and Planning B: Planning and Design*, 34(3):431–445, 2007.
- [128] Roberto G. Trujillo Hsiao-yun Chu. *Buckminster Fuller: A Technocrat for the Counterculture*. 2009.
- [129] M Jeffery Zacks, S Todd Braver, A Margaret Sheridan, I David Donaldson, Z Abraham Snyder, M John Ollinger, L Randy Buckner, and E Marcus Raichle. Human brain activity time-locked to perceptual event boundaries. *In: Nature neuroscience*, pages 651–655, 2001.

- [130] Ullas Gargi. Modeling and clustering of photo capture streams. *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, pages 47–54, November 2003.
- [131] Shûhei Yamamoto, Takuya Nishimura, Yoshiaki Takimoto, Takafumi Inoue, and Hiroyuki Toda. Pbg at the ntcir-13 lifelog-2 lat , lsat , and lest tasks. 2017.
- [132] Ana Garcia del Molino, Joo-Hwee Lim, and Ah-Hwee Tan. Predicting visual context for unsupervised event segmentation in continuous photo-streams. In *ACM Multimedia*, 2018.
- [133] Cathal Gurrin, Alan F. Smeaton, Daragh Byrne, Neil O’Hare, Gareth J. F. Jones, and Noel O’Connor. An examination of a large visual lifelog. In Hang Li, Ting Liu, Wei-Ying Ma, Tetsuya Sakai, Kam-Fai Wong, and Guodong Zhou, editors, *AIRS 2008: Information Retrieval Technology*, pages 537–542, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.
- [134] Jim Gemmell, Lyndsay Williams, Ken Wood, Roger Lueder, and Gordon Bell. Passive capture and ensuing issues for a personal lifetime store. In *Proceedings of the the 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences, CARPE’04*, pages 48–55, New York, NY, USA, 2004. ACM.
- [135] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, November 2004.
- [136] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359, June 2008.
- [137] Sayf A. Majeed, Hafizah Husain, Salina Abdul Samad, and Tariq F. Idbeaa. Mel frequency cepstral coefficients (mfcc) feature extraction enhancement

- in the application of speech recognition: A comparison study. *Journal of Theoretical and Applied Information Technology*, 79(1):38–56, 9 2015.
- [138] Yong Jae Lee and Kristen Grauman. Predicting important objects for egocentric video summarization. *International Journal of Computer Vision*, 114:38–55, 2014.
- [139] Ana Garcia del Molino, Bappaditya Mandal, Liyuan Li, and Lim Joo Hwee. Organizing and retrieving episodic memories from first person view. 07 2015.
- [140] Duc-Tien Dang-Nguyen, Luca Piras, Michael Riegler, Liting Zhou, Mathias Lux, and Cathal Gurrin. Overview of imagecleflifelog 2018: Daily living understanding and lifelog moment retrieval. In *CLEF*, 2018.
- [141] Takayuki Yamada and Kazumasa Takami. Configuration of the system for a fashion coordination service based on clothes life logs. *TENCON 2012 IEEE Region 10 Conference*, pages 1–6, 2012.
- [142] Rashmi "Gupta and Cathal" Gurrin. Considering manual annotations in dynamic segmentation of multimodal lifelog data. In *Proceedings of the ARDUOUS'19-3rd International Workshop on Annotation of useR Data for UbiquitOUS Systems. IEEE European Union*, 2019.
- [143] Rashmi "Gupta and Cathal" Gurrin. Dynamic segmentation and retrieval of daily life activities from multimodal lifelogs. In *Proceedings of the Machine Learning and Data Mining in Pattern Recognition. MLDM 2019*, 2019.
- [144] Wei-Hao Lin and Alexander Hauptmann. Structuring continuous video recordings of everyday life using time-constrained clustering. *Proceedings of SPIE - The International Society for Optical Engineering*, 01 2006.
- [145] Zhe Wang, Matthew D. Hoffman, Perry R. Cook, and Kai Li. Vferret: content-based similarity search tool for continuous archived video. In *CARPE '06*, 2006.

Appendix A

Appendix

A.1 Lifelog Event Segmentation Task Topics

```
# NTCIR13 Lifelog2 Task Event Descriptions
# Topic dataset for the NTCIR12-Lifelog LEST Task
# Last modified: June 28, 2017

<?xml version="1.0" encoding="UTF-8"?>
<topics>
  <topic>
    <id>LEST001</id>
    <type>Event</type>
    <uid>u1, u2</uid>
    <title>Travelling</title>
    <description>The individual is travelling.</description>
    <narrative>Travelling is defined as any event when the
individual is moving from one location to another, regardless of the
locations involved. Any form of transport can be used such as car,
bus, train, boat, bicycle, airplane, etc. Moving in the same
location, such as home or work is also considered travelling. Walking
is not considered travelling. To be considered travelling, the
movement must change location and must last for longer than five
minutes.</narrative>
  </topic>
  <topic>
    <id>LEST002</id>
    <type>Event</type>
    <uid>u1, u2</uid>
    <title>f2f</title>
    <description>The individual has face to face interaction
with people.</description>
    <narrative>Face to face interaction, considered to be any interaction
with one or more people, that can take place in home or in workplace.
Any form of interaction such as business meetings, talking with
colleagues, taking lectures, or having conversation with people in
the home are considered to be face to face interactions. Socialising
outside of the home or workplace is not be considered correct for
this topic. </narrative>
  </topic>
  <topic>
    <id>LEST003</id>
    <type>Event</type>
    <uid>u1, u2</uid>
    <title>Computer</title>
    <description>The individual is using digital
screens.</description>
```

<narrative> Any type of screen is considered correct, such as desktop computer / laptop / tablet / smartphone. Also, reading and writing of content on a computer, tablet or other digital screen is considered correct. For this task, the location such as home or workplace where the users used the digital screens does not matter.</narrative>

</topic>

<topic>

<id>LEST004</id>

<type>Event</type>

<uid>u1, u2</uid>

<title>Cooking</title>

<description>The individual is preparing meals.</description>

<narrative>Preparing meals involves the preparation of food items that occurs before the user, or bystanders, eat the meal. In order to be considered correct, the food preparation process must be visible and can include making tea or coffee or peeling/cutting the vegetables or fruits. The individual themselves must be preparing the food, not simply watching someone else prepare food. The location where the food is prepared does not matter.</narrative>

</topic>

<topic>

<id>LEST005</id>

<type>Event</type>

<uid>u1, u2</uid>

<title>Eating</title>

<description>The individual is eating regular meals.</description>

<narrative>The user must be visibly eating food (where the event lasts at least five minutes). The environment can be at home, workplace, social or in restaurant.</narrative>

</topic>

<topic>

<id>LEST006</id>

<type>Event</type>

<uid>u1, u2</uid>

<title>Children</title>

<description>The individual is physically interacting with children.</description>

<narrative>Any event in which user is physically engaged in taking care of children or playing with children at home or outside environment such as garden / public parks is considered

correct. Playing computer games, watching TV or eating food inside home with children is not considered correct. </narrative>

</topic>

<topic>

<id>LEST007</id>

<type>Event</type>

<uid>u1, u2</uid>

<title>Home</title>

<description>The individual is working in the home.</description>

<narrative>For this task, any event when the user is doing any type of work such as cleaning, gardening, washing utensils, clothes or a car at the home location is considered relevant.</narrative>

</topic>

<topic>

<id>LEST008</id>

<type>Event</type>

<uid>u1, u2</uid>

<title>Relax</title>

<description>The individual is relaxing at home.</description>

<narrative>Relaxing means that the individual is not engaged in any creative or physical activities and must take place at home. Any event that includes watching TV, having a drink or sitting in the backyard is considered correct. Relaxing events are orthogonal to other events in this task.</narrative>

</topic>

<topic>

<id>LEST009</id>

<type>Event</type>

<uid>u1, u2</uid>

<title>Paper</title>

<description>The individual is reading content on a physical medium.</description>

<narrative>Any event when the user is reading a book, newspaper, map, or other physical printed content. Reading of content on a computer, tablet or other digital screen is not considered correct. </narrative>

</topic>

<topic>

<id>LEST010</id>

<type>Event</type>

```

        <uid>u1, u2</uid>
        <title>Socialising</title>
        <description>The individual is socialising.</description>
        <narrative>Socialising is defined as any collection of
moments in which the user is engaged in face to face conversation
with other people in a non-work and non-home environment.</narrative>
    </topic>
    <topic>
        <id>LEST011</id>
        <type>Event</type>
        <uid>u1, u2</uid>
        <title>Pray</title>
        <description>The individual is praying / worshipping or
meditating.</description>
        <narrative>Any event when the individual is praying,
worshipping or meditating at home or outside the home. This can take
place indoors (e.g. in a church) or outdoors (e.g. meditating in a
park)</narrative>
    </topic>
    <topic>
        <id>LEST012</id>
        <type>Event</type>
        <uid>u1, u2</uid>
        <title>Shopping</title>
        <description>The individual is shopping in a physical
shop.</description>
        <narrative>Any event when the individual is shopping in a
physical shop, buying groceries in supermarkets or buying any type of
physical accessory such as shoes, electronic items, hardware items,
clothes or kitchen appliances, etc. is considered to be correct.
Shopping online by using digital technology is not considered to be
appropriate for this task.</narrative>
    </topic>
    <topic>
        <id>LEST013</id>
        <type>Event</type>
        <uid>u1, u2</uid>
        <title>Gaming</title>
        <description>The individual is playing computer
games.</description>
        <narrative>For this task, any event when the individual is
playing computer games at home or non-home environment is considered
correct. Playing games on desktop computers, tablet or laptops are

```

considered to be relevant. Any event when user is playing video games on television or mobile phone, or playing outdoor sports such as football, rugby or any physical activity is not considered to be relevant. </narrative>

</topic>

<topic>

<id>LEST014</id>

<type>Event</type>

<uid>u1, u2</uid>

<title>Physical Activities</title>

<description>The individual is engaged in any physical activities or outdoor sports at non-home and non-work environment.</description>

<narrative>We define physical activity as any event when individual is engaged in any physical activity such as exercise, walking, cycling, running, rowing or playing outdoor sports (football, hockey or rugby, etc). Any such activity must last for longer than five minutes. The exercise can take place at any location such as a gym, garden, once it occurs outside the home. </narrative>

</topic>

<topic>

<id>LEST015</id>

<type>Event</type>

<uid>u1, u2</uid>

<title>Creative endeavours</title>

<description>The individual is doing any creative artistic work.</description>

<narrative> Any event when individual is writing on a physical medium of paper using pen/pencil or writing using digital technology such as laptop, desktop computer, tablets or smartphones. Events such as painting, sketching or drawing on painting boards or paper is considered correct. Painting or sketching using digital technology such as a computer, laptop, tablet or smartphone is also considered relevant for this task. Playing with and form of musical instrument is also to be considered relevant for this task. </narrative>

</topic>

<topic>

<id>LEST016</id>

<type>Event</type>

<uid>u1, u2</uid>

<title>Others</title>

<description>The individual is not engaged with above discussed activities.</description>

<narrative> For this task, any other event or activity that is not represented by the fifteen labels above are considered to be relevant. </narrative>

</topic>

</topics>

A.2 Ontology to Annotate Lifelog Data

NTCIR-13 Lifelog-2 Annotation Task (ver 0.2)

There are two types of annotation that we are interested in exploring in lifelog data for the LAT task of NTCIR-13 Lifelog-2. They are the activities / facets of daily life and the environmental settings / contexts in which these activities take place. For the LAT task, we will choose a subset of ten of these activities and settings for evaluation.

Activities / Facets of Life Activity

- Commuting (to/from work or other work-related venue)
- Travelling (to a destination other than work, home)
- Preparing meals (include making tea or coffee)
- Eating/drinking
 - Eating food
 - Drinking drinks
- Taking care of children / playing with children
- Sleeping
- Praying / worshipping / meditating
- Socialising / casual conversation
- Relaxing / meditation
- Reading
 - reading a book/paper
 - reading digital content
- Gardening
- Shopping
 - retail shopping and purchasing
 - browsing for items in store
 - online shopping
- Work meeting/interaction (one or more people)
 - face-to-face
 - online meeting
- Watching (TV)
- Playing computer games
- Using desktop computer / laptop computer
- Using mobile device / tablet
- Physical activities / sports
 - walking
 - running
 - cycling
 - other sports/exercise
- Searching / Information Seeking
 - WWW
 - media seeking
 - files

- locations (Map)
- Organising things
 - papers
 - boxes
 - other
- Packing bags/suitcases
- Cleaning
 - home cleaning
 - vacuuming
 - dish washing
- Hygiene & make-up
 - hands
 - face
 - teeth
 - make-up
 - other personal hygiene
- Artistic endeavours
 - writing
 - drawing / painting
 - playing music instruments

Settings/Context (the physical environment)

- In an office environment
- In a home
 - in a kitchen
 - in a living room
 - in a bedroom
 - in a garden
- In a publicly-accessible building
 - in a shop / mall
 - in a restaurant
 - in a hospital
 - in a school / college
 - in a church
 - at a tourist site
 - other
- In an open space
 - street
 - public green-space
 - by the sea
- In a vehicle (as a driver or passenger - inside)
 - car (driver)
 - car (passenger)

A.3 LADT - Lifelog Activity Detection Task Topics

The aim of this subtask is to develop new approaches to the annotation of multimodal lifelog data in terms of activities of daily living. An ontology of important lifelog activities of daily living act as the ground truth for the development effort (see below). The task requires the development of automated approaches to annotating these concepts from multimodal lifelog data. Both image content as well as provided metadata and external evidence sources can be used to generate the activity annotations. The submission is in terms of minutes where every minute is annotated with the desired activity. There are sixteen activities identified below:

To reduce the level of subjectivity in the manual judgement process, the manual segmentation is achieved by segmentation the 90 days of data into fifteen distinct episode types, as follows (label: description):

- **traveling**: travelling (car, bus, boat, airplane, train, etc)
- **f2f interacting**: face-to-face interaction with people at home or in the workplace (excluding social interactions)
- **using a computer**: using desktop computer / laptop / tablet / smartphone
- **cooking**: preparing meals (include making tea or coffee) at any location
- **eating**: eating meals in any location, but not including moments when drinking alone.
- **time with children**: taking care of children / playing with children
- **house working**: working in the home (e.g. cleaning, gardening)
- **relaxing**: relaxing at home (e.g. TV, having a drink)
- **reading**: reading any form of paper
- **socialising**: socialising outside the home or office
- **praying**: praying / worshipping / meditating
- **shopping**: shopping in a physical shop (not online)
- **gaming**: playing computer games
- **physical activities**: physical activities / sports (walking, playing sports, cycling, rowing, etc)
- **creative activities**: creative endeavours (writing, art, music)
- **other activities**: any other activity not represented by the fifteen labels above.

A.4 Appendix to Publications

- **Peer-reviewed:** Duc Tien Dang Nguyen, Liting Zhou, **Rashmi Gupta**, Michael Riegler, Cathal Gurrin (2017) Building a disclosed lifelog dataset : challenges, principles and processes. In: Content-based Multimedia Indexing, 19-21 Jun 2017, Florence, Italy. ISBN 978-1-4503-5333-5.
- **Peer-reviewed:** Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, **Rashmi Gupta**, Rami Albatal, Duc-Tien Dang-Nguyen (2017) Overview of NTCIR-13 Lifelog-2 Task. In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies, December 5-8, 2017, Tokyo, Japan.

Contribution of Dissertation writer: The dissertation writer was a co-organisier of the LEST

(lifelog event segmentation task) at NTCIR evaluation workshop and she needed to provide her opinion on what sources of data would be needed and proved to be beneficial in the LEST task for the target analysis.

Further to this, as the research team was dealing with lifelog data which is sensitive in nature, the main responsibility of dissertation writer was to ensure the willingness of the users to share data to the third party organisations and to the public for the research purposes. However, finding people who were willing to donate (months or years) of their personal data was a challenge. Also privacy and data security, which has implications for both the individual and society as a whole, were also important issues that need to be considered. This raised the challenge of how to filter out the private information in order to protect the lifeloggers as well as individuals that appear in the lifelogs. In order to resolve this, the dissertation writer contributed to engage in a user study to understand the variance of privacy preference across participants while sharing their lifelog data to public or third party organisations that helped the research team members to understand the users privacy concerns and what data governance guidelines need to be provided to the users. Thus, the research team used the findings and potential data governance guidelines from this user study that provided the insights to the researchers/third-party organisations building digital trust and best practices for data sharing. Hence, the privacy-aware lifelog test collection that supports comparative and repeatable experimentation in this research was generated.

- **Journal:** Rashmi Gupta, Martin Crane, Cathal Gurrin (2020). Considerations on Privacy in the era of Digitally Logged Lives. In: Online Information Review (OIR), December 2020.

Contribution of Dissertation writer: To validate our first and second fundamental research questions proposed in Chapter 1 and the above mentioned study motivated me to propose a user study where I tend to observe the privacy concerns of the users participating in lifelogging domain and how we can resolve these concerns to make lifelog data publicly available to the researcher for further data analysis.

- **Peer-reviewed:** Rashmi Gupta (2018) Considering documents in lifelog information retrieval. In: International Conference on Multimedia Retrieval ACM (ICMR 2018), 11-14 June 2018, Yokohama, Japan. ISBN 978-1-4503-5046-4.

Contribution of Dissertation writer: In this publication, the dissertation writer was focused on resolving the third research question proposed in Chapter 1 and proposed a hierarchical ontology to define different units of retrieval in lifelog domain and found the best suitable unit of retrieval for her research.

- **Peer-reviewed:** Rashmi Gupta, Cathal Gurrin (2018). Approaches for event segmentation of visual lifelog data. In: 24th International Conference on Multimedia Modeling (MMM 2018), 5-7 Feb, 2018, Bangkok, Thailand.

Contribution of Dissertation writer: In this publication, the dissertation writer continued with the more scientific research question (i.e. RQ4 explained in Chapter 1) and built data analysis model to segment privacy-aware visual lifelog data into events.

- **Peer-reviewed:** Rashmi Gupta, Cathal Gurrin (2019). Dynamic Segmentation and Retrieval of Daily Life Activities from Multimodal Lifelogs. In Press: Perner P. (eds) Machine Learning and Data Mining in Pattern Recognition (MLDM 2019). Lecture Notes in Computer Science. Springer International Publishing.

Contribution of Dissertation writer: In this publication, the dissertation writer continued to propose a dynamic segmentation model to segment continuous multimodal lifelog data at retrieval time as per user information needs (i.e. RQ5 explained in Chapter 1) and built an ad-hoc search engine for users.

- **Workshop:** Rashmi Gupta, Cathal Gurrin (2019). Manual Annotations in Dynamic Segmentation of Multimodal Lifelog Data. In: 3rd International Workshop on Annotation of user Data for Ubiquitous Systems (ARDUOUS 2019), March 11-15, Kyoto, Japan.

Contribution of Dissertation writer: In this publication, the dissertation writer extended the previous study of the proposed dynamic segmentation model and introduced a new idea of observing the impact of manual annotations in the overall accuracy of the system (i.e. RQ5 explained in Chapter 1).

- **Peer-reviewed: Rashmi Gupta, Cathal Gurrin (2019).** Activity Recognition and Segmentation Approaches to Multimodal Lifelog Data. In: International Conference on Content-Based Multimedia Indexing (CBMI 2019), Sept 4-6, Dublin, Ireland.

Contribution of Dissertation writer: In this publication, the dissertation writer extended the previous study of the proposed dynamic segmentation model and focused on segmenting multimodal lifelog data based on activities as basic unit of retrieval (i.e. authors found “activity” is the best unit of retrieval in this research). Authors compared both static (indexing-time) and dynamic (retrieval-time) segmentation approaches and identified the best segmentation approach that holds the validation of our proposed hypothesis as well as resolved the fifth research question.

- **Peer-reviewed: Aaron Duane, Rashmi Gupta, Liting Zhou, Cathal Gurrin (2016).** In: Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, June 7-10, 2016 Tokyo Japan.

Contribution of Dissertation writer: In this publication, the dissertation writer was engaged in LIT (Lifelog Insight Task) organised by NTCIR evaluation workshop in 2016. The main contribution of the dissertation writer in this work was to do descriptive analysis in observing the regularity in daily activities of an individual such as early morning commute, usage of public transportation, maintaining healthy lifestyle, diet monitoring, and usage of digital screen, which provided a very good knowledge to understand lifelog data to dissertation writer in her early stage of PhD.