DCU at the NTCIR-13 Lifelog-2 Task

Liting Zhou
Dublin City University
zhou.liting2@mail.dcu.ie

Aaron Duane
Dublin City University
aaron.duane@mail.dcu.ie

Duc-Tien Dang-Nguyen
Dublin City University
duc-tien.dang-nguyen@dcu.ie

Cathal Gurrin
Dublin City University
cgurrin@computing.dcu.ie

ABSTRACT

In this work, we outline the submissions of Dublin City University (DCU) team, the organisers, to the NTCIR-13 Lifelog-2 Task. We submitted runs to the Lifelog Semantics Access (LSAT) and the Lifelog Insight (LIT) sub-tasks.

Team Name

DCU

Subtasks

Lifelog - 2 (LSAT, LIT)

Keywords

Lifelogging, Search Engine, Personal Life Archive, VR, Multimedia Retrieval

1. INTRODUCTION

Thanks to the advance of technologies, wearable devices and quantified-self applications [14] are becoming more popular in the recent years. Every moment and aspect of our lives can now can be easily stored into personal live archives, often called lifelogs [11]. Such lifelogs can capture a detailed trace of human activities [6, 16], from meeting friends to consuming foods [1], and are increasingly interesting for the research community for ethnographic [15] and memory studies [12, 2, 8], as well as companies. Discover insights from personal data is a trend now. Collecting and storing such kind of data for research, however, is non-trivial with many challenges from privacy and other concerns [5]. Moreover, getting insights from the collected data and finding new information by connecting different types of data requires a significant investment of effort in analysing, categorising and accessing these huge volumes of data in an efficient manner. This is, as of yet an unsolved challenge, and consequently several benchmarking campaigns and workshops have been organised in recent years, such as the NTCIR-13 - Lifelog 2 task [10], the ImageCLEF2017lifelog [4] in 2017, and the LTA workshops [9] in recent years, to offer forums, common data, and tasks for researchers to study.

In this paper, we report on the DCU activities for the latest such benchmarking exercise, the NTCIR-13 - Lifelog 2 task [10] which aims to advance the state-of-the-art research in lifelogging as an application of information retrieval. In this paper, we present our approaches to tackle the Lifelog Semantics Access (LSAT) and Lifelog Insight (LIT) subtasks at NTCIR-13 - Lifelog 2 task. The first subtask, LSAT, aims at solving the retrieval problems by asking participants to retrieve a number of specific moments in a lifelogger's life. Moments are defined as semantic events, or activities that happened throughout the day. The second subtask, LIT, is aiming at exploring access methods that provide insights into the lifelogger's life, across a certain number of named dimensions. It follows the idea of the Quantified Self movement that focuses on the visualization of knowledge mined from self-tracking data to provide "self-knowledge through numbers". Participants are requested to generate new types of visualisations and insights about the life of the lifeloggers by generating a themed diary or themed insights related to some target topics.

For LSAT, we report on the baseline retrieval method proposed in [17], which (as task organisers) we made available for all participants to use. This baseline search engine provided retrieval facilities over the lifelog dataset for the LSAT subtask and it indexed all available metadata provided with the test collection. The aim of the baseline search engine was to provide a point of comparison for other participants, and to lower the barriers-to-entry for new participating teams by providing basic search functionality via an API. The basic idea is to turn all metadata into feature vectors and then index those features for the retrieval. To address the LIT subtask, we create a matrix of concepts related to the theme of the selected topics and formulate them into a batch filter query which would be run in real-time when prompted by the user. After the batch filter query is performed, the user is presented with a virtual wall of images which correspond to the chosen topic. This wall can be scrolled, navigated and resized in any of the four virtual reality interaction methodologies, which will be outlined later in this paper.

The remainder of this paper is organized as follows, first we present our work on the LSAT subtask. This is followed by a detailed description of our approach on the LIT subtask. Finally, we draw some conclusions and future work.

2. LSAT SEARCH ENGINE

The main idea behind this baseline search engine is first to provide a starting point for researchers in the area, as well as a documented system for comparative analysis. The baseline search engine extracted textual from the raw data for higher level analysis. Then, it indexes all the extracted features and hierarchically organises them which allow queries to be processed via an API/interface.

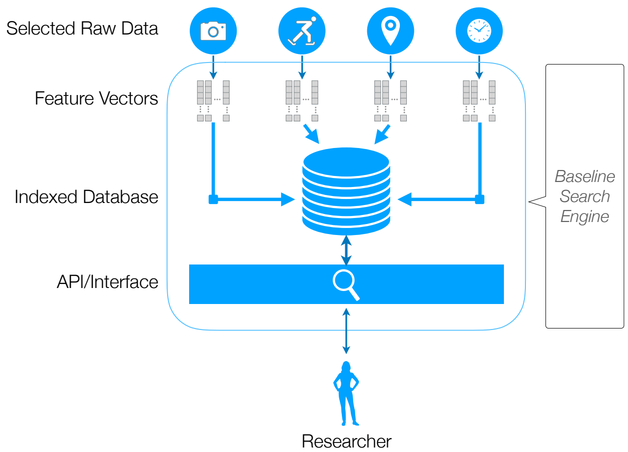


Figure 1: The baseline search engine architecture

2.1 Search Engine Architecture

Figure 1 summarises how we designed the baseline search engine system [17], as follows: from the provided raw data, we extract locations, visual concepts, time, and activities and transformed this raw data into indexable feature vectors. These feature vectors are then indexed and hierarchically organized. Finally, a user or other system can define a faceted query and retrieve ranked moments via the interface. The basic architecture is shown in Figure 1.

2.2 Data Organisation and Retrieval Process

To organise and index the data, we follow the study in [17] by arranging features as chronological order, and use the minute as the basic units. Building up from these basic units, we organize the data at higher level which can be turned into more useful information, the minutes are hierarchically grouped into event nodes (typically, in a full day, a person encounters anything upwards of 20 individual events, with each lasting (on average) 30 minutes [7, 13]), then ultimately leading to larger units such as days and multi-day events (e.g. holidays).

The retrieval is then simply done by ranking moments (in this task, images) that matched with the queried criteria. The challenge here is how to turn a query into specific criteria. It can be automatically done by considering any word in the topic as the queried concepts and then searching for all images that contain those concepts or by fine-tunning the query manually by a human (in the loop), i.e., the researcher (we) will read the topic and "translate" it into the search criteria. For this task, we applied both methods and the results will be reported in the result section. For example, with the query:

`"Find the moments when I was outside at sunset."`

we follow the study in [18] and "translate" that query into specific required pieces of information, as follows:

- `User = {u1}`,
- `Concepts = {sunset}`,
- `Activity = {walking, standing}`,
- `time = {17:00 - 19:00}`
- `Location = {outside}`

2.3 Ranking

To refine the results, i.e., to increase the precision of the top retrieved images, we use a hierarchical agglomerative clustering algorithm (see [3]) to group similar images into the same cluster based on all of their features. The clusters are then sorted based on the number of images, in decreasing rank order. Finally, we produce the retrieval by selecting representative images from the clusters by choosing the images closest to the center of each cluster.

2.4 Results

We applied the baseline search engine to this task in two submissions: the first one is a fully automatic approach by taking all words in the query as the searching criteria. In the second submission, we manually "translate" the topic into criteria.

With automatic method, the official result is obtained at 0.098. With the human-in-the-loop method in the second submission, we got a score of 0.329. Some examples of this submission can be seen in Figure 2.

Looking into details, the precision and recall at cut-off point at 10 we got are 0.48 and 0.11, respectively. At some topic, the results reaching the "golden results", i.e., the score at the ground-truth, for example at topic LSAT003 (`"Find the moment when I was visiting a castle at night"`) we got 1.00 and 0.92 for precision and recall, respectively, while at some topics, for example LSAT019 (`"Find the moment when I was painting the walls at home"`), the results were all zeros at cut-off point at 10. It shown that further searches are required to improve the retrieval on this kind of data.

We also built and made this baseline search engine open access at: `http://search-lifelog.computing.dcu.ie`, which are now serving as the basic tool for other participants at NTCIR-13 - Lifelog 2 task.

3. LIT: VIRTUAL REALITY LIFELOG EXPLORATION

One of the rarely considered research challenges in lifelogging is how to support user access from non-standard devices. For the LIT task, rather than developing a suite of themed diaries, we developed a lifelog interrogation tool that operated in a Virtual Reality (VR) environment. It is our conjecture that the multidimensional nature of VR systems presents opportunities for effective visual lifelog exploration, due to the nature of lifelog datasets which often contain multivariate data often requiring temporal alignment from many heterogeneous sources. The virtual reality lifelog exploration prototype developed for the LIT subtask was created using the Unity game engine, SteamVR and the HTC Vive. The prototype provides the user with four topic labels, matching four of the five LIT subtask themes (Diet, Social, Exercise, Location).

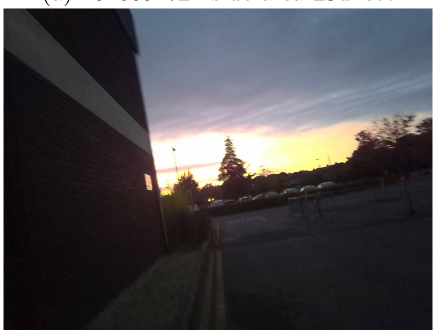
Additionally, there were four different types of interaction method provided to the user, that were inspired by the most prevalent mechanisms observed in the industry at the time of development (in mid-late 2017).

The first and most commonly observed interaction type is via a point and click style interface, similar to how one might interact with a television using a remote control. In our implementation for the LIT subtask, the user simply points their Vive controller at an element of interest and a beam will appear which enables them to interact with that element

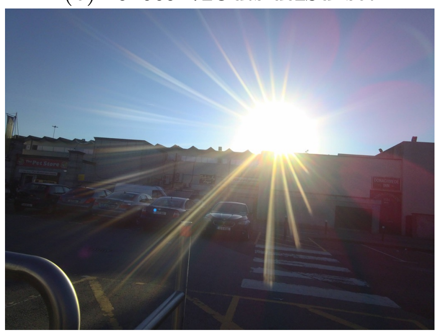
(a) 20160917_Inside a car_Sunset

(b) 20160917_Outside_Sunset

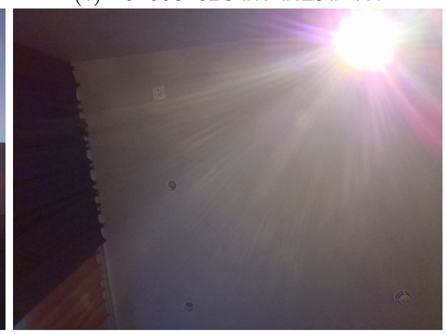
(c) 20160825_Outside_Sunset



(d) 20160913_Outside_Sunset



(e) 20161005_Inside a car_Sunset



(f) 20160905_Home_Lamplight

(g) 20160922_Outside_Sunset

(h) 20160823_Outside_Sunset

(i)20161011_Outside_Sunset

Figure 2: Examples of the results retrieved by the proposed baseline search engine for the query "Find the moment when I was outside at sunset." According to this result, (a) and (f) are non-relevant while the other seven images are relevant.

(see Figure 3) and to select which topic label is of interest. In this approach, the menu elements are presented upright and parallel to the user. The advantage of this approach is that it is intuitive for the average user but the disadvantage is the menu elements must appear in front of the user and can obscure other content in the virtual environment.

The second interaction type works similarly to the first approach, in that it works via a point and click interface using a beam, but differs in that the menu elements being interacted with are presented horizontally on the floor at the user's feet. This approach enables the menu to exist without obscuring other content in the virtual environment but is not as intuitive to the average user.

The third interaction approach involves attaching a drumstick-style object to the top of one Vive controller then attaching the target menu to the opposing Vive controller. The user then interacts with the menu elements by tapping the drumstick tip against what they want to interact with, as an artist uses an artist's easel to select paint.

The fourth and final interaction approach relies on adding the drumstick-style objects to both controllers instead of just one. Then the menu elements are presented in front of the user and tiled at a slight angle, similar to an elevated keyboard or desk. The user may then interact with any desired menu elements by tapping on it with either controller's drumstick.

To address the topics outlined in the lifelog insight task within the virtual environment, we created a matrix of concepts related to the theme of the selected topic. For example, to explore the topic of 'diet', we combined concept terms such as 'food', 'drink', 'eating', 'coffee', etc. and formulated them into a batch filter query which would be run in real-time when prompted by the user. After the batch filter query is performed, the user is presented with a virtual wall of temporally organised lifelog images which correspond to the chosen topic (see Figure 4). This wall can be scrolled, navigated and browsed using any of the four virtual reality interaction methodologies outlined above. In addition

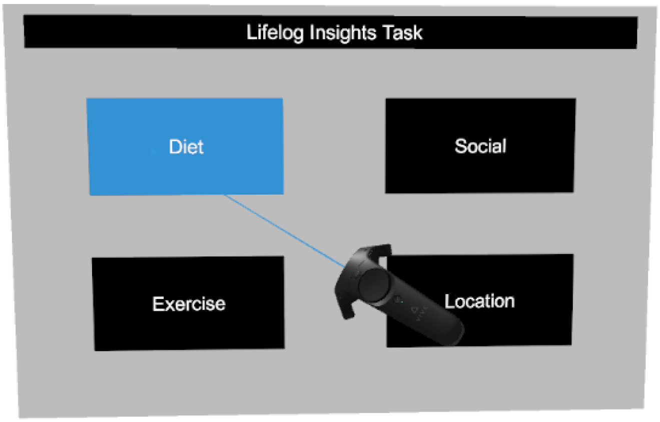


Figure 3: Selection Menu

Figure 4: Memory Wall

to this, any image can be selected to observe its metadata via a virtual overlay. It is our proposal that such a system can immediately give a user a visual representation of their activities and that it can act as a VR-based visual diary to provide the user with a high-level overview of their past activities, while at the same time, allowing them to retrieve lower-level details from any image/moment of interest.

4. CONCLUSIONS AND FUTURE WORK

The potential for personal life archives is enormous. We firmly believe that personal data research will be the future and exploiting the personal life archives will positively impact on everyone who uses the technology. In this paper, we introduced methods for retrieving basic moments from the lifelog data in a reliable and efficient manner by exploiting the baseline search engine. We also addressed different way of interactions using VR which allow lifelogger gain insights from their lifelogs. In the future work, we plan to enrich, and extend the search engine with more agile and advanced solutions, aiming at giving better information for higher level of insight and query engines. These results will also be served as the input for further analysis, which allows lifeloggers discover insights from their life by using the novel VR interactions.

5. ACKNOWLEDGEMENT

This publication has emanated from research supported in party by research grants from Science Foundation Ireland under grant number SFI/12/RC/2289 and Irish Research Council (IRC) under Grant Number GOIPG/2016/741.

6. REFERENCES

[1] K. Aizawa. Multimedia foodlog: Diverse applications from self-monitoring to social contributions. *ITE Transactions on Media Technology and Applications*, 1(3):214–219, 2013.

[2] E. Berry, A. Hampshire, J. Rowe, S. Hodges, N. Kapur, P. Watson, G. B. G. Smyth, K. Wood, and A. M. Owen. The neural basis of effective memory therapy in a patient with limbic encephalitis. *Neurology, Neurosurgery, and Psychiatry with Practical Neurology*, 80(3):582–601, 2009.

[3] D.-T. Dang-Nguyen, L. Piras, G. Giacinto, G. Boato, and F. G. De Natale. Multimodal retrieval with diversification and relevance feedback for tourist attraction images. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2017. Accepted.

[4] D.-T. Dang-Nguyen, L. Piras, M. Riegler, G. Boato, L. Zhou, and C. Gurrin. Overview of ImageCLEFlifelog 2017: Lifelog Retrieval and Summarization. In *CLEF2017 Working Notes*, CEUR Workshop Proceedings, Dublin, Ireland, 2017. CEUR-WS.org <http://ceur-ws.org>.

[5] D.-T. Dang-Nguyen, L. Zhou, R. Gupta, M. Riegler, and C. Gurrin. Building a Disclosed Lifelog Dataset: Challenges, Principles and Processes. In *Content-Based Multimedia Indexing (CBMI)*, 2017.

[6] M. Dodge and R. Kitchin. “Outlines of a world coming into existence”: Pervasive computing and the ethics of forgetting. *Environment and Planning B*, 34(3):431–445, 2007.

[7] A. Doherty and A. Smeaton. Automatically segmenting lifelog data into events. In *Ninth International Workshop on Image Analysis for Multimedia Interactive Services*, pages 20–23. IEEE, 2008.

[8] A. R. Doherty, K. Pauly-Takacs, N. Caprani, C. Gurrin, C. J. A. Moulin, N. E. OÊijConnor, and A. F. Smeaton. Experiences of aiding autobiographical memory using the sensecam. *Human-Computer Interaction*, 27(1-2):151–174, 2012.

[9] C. Gurrin, X. Giro-i Nieto, P. Radeva, M. Dimiccoli, D.-T. Dang-Nguyen, and H. Joho. LTA 17: The second Workshop on Lifelogging Tools and Applications. In *Proceedings of the 2nd Workshop on Lifelogging Tools and Applications*. ACM, 2017.

[10] C. Gurrin, H. Joho, F. Hopfgartner, L. Zhou, R. Gupta, R. Albatal, and D.-T. Dang-Nguyen. Overview of NTCIR-13 Lifelog-2 Task, 2017.

[11] C. Gurrin, A. F. Smeaton, and A. R. Doherty. Lifelogging: Personal big data. *Foundations and Trends in Information Retrieval*, 8(1):1–125, 2014.

[12] S. Hodges, E. Berry, and K. Wood. Sensecam: A wearable camera that stimulates and rehabilitates

autobiographical memory. *Memory*, 7(19):685–696, 2011.

[13] D. Kahneman, A. B. Krueger, D. A. Schkade, N. Schwarz, and A. A. Stone. A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306:1776–1780, 2004.

[14] J. Meyer, S. Simske, K. A. Siek, C. G. Gurrin, and H. Hermens. Beyond quantified self: Data for wellbeing. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, pages 95–98, New York, NY, USA, 2014. ACM.

[15] J. Rieken, E. Garcia-Sanchez, M. P. Trujillo, and D. Bear. Digital ethnography and the social dimension of introspection: An empirical study in two colombian schools. *Integrative Psychological and Behavioral Science*, 49(2):253–274, Jun 2015.

[16] D. Tancharoen, T. Yamasaki, and K. Aizawa. Practical life log video indexing based on content and context. In *Multimedia Content Analysis, Management, and Retieval ï£¡ In Proceedings of SPIE-IST Electronic Imaging*, 2006.

[17] L. Zhou, D.-T. Dang-Nguyen, and C. Gurrin. A baseline search engine for personal life archives. In *Proceedings of the 2nd Workshop on Lifelogging Tools and Applications*, LTA '17, pages 21–24, New York, NY, USA, 2017. ACM.

[18] L. Zhou, L. Piras, M. Rieger, G. Boato, D.-T. Dang-Nguyen, and C. Gurrin. Organizer Team at ImageCLEFlifelog 2017: Baseline Approaches for Lifelog Retrieval and Summarization. *CLEF working notes, CEUR*, September 11-14 2017.