An Interactive Lifelog Retrieval System for Activities of Daily Living Understanding

Liting Zhou[1], Luca Piras[2], Michael Riegler[3], Mathias Lux[4], Duc-Tien Dang-Nguyen[1], and Cathal Gurrin[1]

[1] Dublin City University
zhou.liting2@mail.dcu.ie, {duc-tien.dang-nguyen, cathal.gurrin}@dcu.ie
[2] Pluribus One & University of Cagliari
luca.piras@diee.unica.it
[3] Simula Research Laboratory
michael@simula.no
[4] ITEC, Klagenfurt University
mlux@itec.aau.at

Abstract. This paper describes the participation of the Organizer Team in the ImageCLEFlifelog 2018 Daily Living Understanding and Lifelog Moment Retrieval. In this paper, we propose how to exploit LIFER, an interactive lifelog search engine to solve the two tasks: Lifelog Moment Retrieval and Activities of Daily Living Understanding. We propose approaches for both baseline, which aim to provide a reference system for other approaches, and human-in-the-loop, which advance the baseline results.

1 Introduction

A new trend in multimedia research is the generation of personalized archives that store rich detail of your life experience using various modalities such as videos, images, text or sensor data. These logs are commonly refereed to as lifelogs [12]. Lifelogs typically contain details of your life experience, such as consumed food, visited places and many more. Such rich archives hold a lot of potential not just for research but also for the users themselves.

Lifelogging poses many challenging research questions [5], such as how to make this rich data searchable, how to extract meaningful information and how to summarize data, etc. Regarding this, a couple of initiatives have been organised in the last few years, for example NTCIR-13 [9], ImageCLEFlifelog2018 [4] at ImageCLEF 2018 [16] which the goal is to bring researchers in different domains to solve the challenges in the novel research field.

In this paper we describe our solution to the 2018 Image CLEF [16] Lifelog Task [4]. For our approach we exploit the fact that Lifelogs are usually chronologically organized. Hence, moments that belong to the same activity or event are likely to be very similar. By performing similarity or near duplicate detection, we can group moments based on time and concepts. Tackling the problem with this angle transforms the image retrieval challenge into an image segment retrieval challenge.

Utilizing time and concepts comes with the advantage that boundaries between events are easily identifiable. This saves both processing time and computation power [6]. In addition to this we also remove images that do not contain much information and would rather add noise to the analysis (blurry, one object, etc.). In our past work, this was estimated to be in the region of 40% of all images [11]. Images retrieved by our method are then clustered for used in other tasks, for example, summarization, classification or search [3].

The paper is organized as follows: Firstly, we provide an overview of related work in the field. After that we give a detailed description of LIFER, the interactive system that our approach is based on, which is followed by a methodology of how to exploit the system. We then show the results obtained from the official competition and finally, we discuss the solutions, the results, and conclude the paper.

2 Related Work

In general there exits an ever increasing body of research aimed at solving different aspects of the overall lifelogging information access challenge, ranging from computer vision [20] to multidimensional visualization of lifelogging data [13].

In context of our approach, several related works exist. A common practice for image segmentation based on time data is heuristic splitting [18]. Another often used technique is based on utilizing thresholds on the distances between images related to the content [2]. Apart from these supervised approaches also unsupervised methods exist [1].

As with other fields of research, information retrieval currently relies heavily on deep learning approaches [14, 19]. Current work focuses on retrieval results that represent relevant and diverse samples of the archives [15]. This trend is also emerging in lifelogging. Fan et al. [7] propose a deep learning approach to perform image caption and summarization for lifelogging datasets. Nevertheless, deep learning within lifelogging still struggles with some challenges. One of them is that multi-modal deep learning is not well researched [17] yet. Therefore approaches that rely on traditional methods still perform better. For example [3] where the authors rely on relevance feedback to retrieve relevant and diverse results and at the same time keep the number of iterations low.

3 LIFER: An Interactive Lifelog Search Engine

Our proposed solutions for the two tasks is to exploit the baseline system, LIFER [22], which is an interactive engine for lifelog retrieval. LIFER is improved upon an existing baseline search engine [21], which was developed to provide a starting point for researchers engaged in collaborative benchmarking exercises, such as NTCIR [9] and this ImageCLEFlifelog2018 [4] tasks. It was also used for the LSC@ICMR 2018 [10] competition and got a reasonable result.

In this section, we will introduce how we used the LIFER system to address the two tasks and list the approaches we used for retrieval images information.

LIFER uses the core search engine of [21] which offers a platform which can be used to search for images that match with some criteria. The retrieved moments (represented by an image for each moment) are then presented to the user in temporal order. Since the collection was small, this temporal order is unlikely to be too large for fast human browsing and selection. This interactive system helps user to retrieve results in a faster and reliable way, which helps to solve both two tasks. The detailed operation will be described in Section 4.

LIFER is built based on the six sources of information which was extracted from ImageCLEF dataset offered.

- Time. The most basic unit of data in the dataset, time gave us the possibility of including more semantic concepts, such as days of the week, weekday/weekend, times of the day, etc. In the LIFER system, we consider the unit of time as minute, i.e., each image is attached to a minute. These time is extracted (and linked to the image) directly from the provided data.
- Locations. Semantic location were provided in the dataset which provided localised names for all locations visited. For example 'The Helix', 'Dunnes stores', 'Dublin City University' and so on.
- Visual Concepts. visual concepts extracted by Microsoft API[8] was provided, which accompany with each image. These visual concepts were indexed in our lifelog retrieval system. Visual concepts describe the content of the lifelog images included in the dataset. Each image has one (or more) concepts identified and tagged. The concepts (in text form) were indexed.
- User Activities. The physical activities of the user (e.g. walking, sitting, running, etc.) were indexed as additional search terms.
- Biometrics. The biometrics of the user were also indexed as semantic labels. These included the Galvanic Skin Response (stressed/excited, relaxed) which can be considered to be a correlate of stress or excitement levels, and the level of physical activity (exertion / resting) as identified from the heart rate.
- Music. A log of the music listing history of the lifelogger was included in the collection and we considered that it could be an important aspect of some topics. The song_name and song_artist are two options which are used to search results.

These six sources of information are instantiated in the user interface as facets of a user query, as shown in Figure 1.

The Interface of LIFER is shown in Figure 2. The upper section of the interface is the query-panel in which the faceted queries are created. Below that is the main part of the interface which is where the selected lifelog images are displayed in temporal sequence.

In the query-panel, the search facets are shown. The facets are directly related to the indexed data (see the six sources of information above). Upon submission of a faceted query, the system returns a temporally organised listing of potentially relevant images. In the first version of LIFER, the query facets are combined

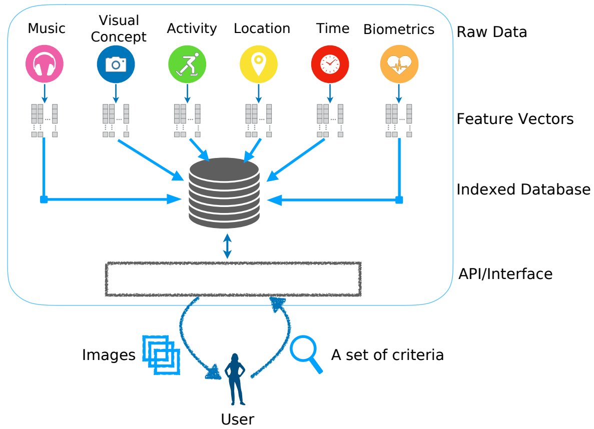

Fig. 1. Schema of LIFER, the proposed interactive lifelog search engine.

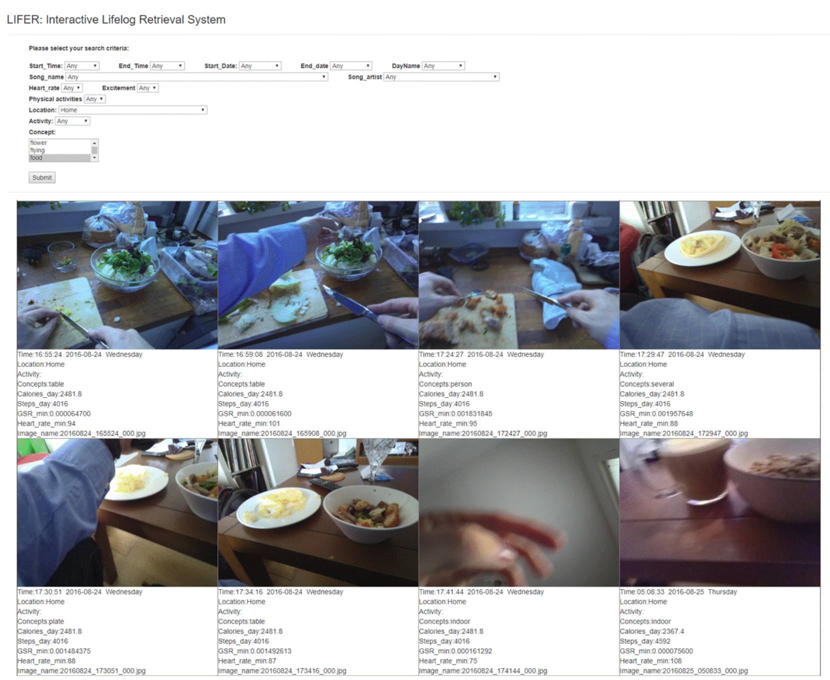

Fig. 2. The Interface of LIFER (http://search-lifelog.computing.dcu.ie) with an example on the results for solving Topic 1 of the LMRT task.

in an AND boolean manner. This can be changed on a per-topic basis, but does not form part of the interface at present.

Table 1. Submitted Runs for ADLT task.

RunID	Name	Similarity	Notes
ADLT Run 1	Baseline	0.816	Search by main terms and Human filtering
ADLT Run 2*	Baseline	0.456	Search by relevant terms and Human filtering
ADLT Run 3*	Baseline	0.344	Search by relevant terms and Human filtering
ADLT Run 4*	Baseline	0.481	Search by relevant terms and Human filtering
ADLT Run 5*	Baseline	0.485	Search by relevant terms and Human filtering

* These runs were submitted after the competition.

Table 2. Submitted Runs for LMRT.

RunID	Name	F1@10	Notes
LMRT Run 1	Baseline	0.077	fully automatic without ranking
LMRT Run 2	Baseline	0.131	fully automatic with ranking
LMRT Run 3*	Baseline	**0.407**	Search by main terms and Human filtering
LMRT Run 4*	Baseline	0.378	Search by relevant terms and Human filtering
LMRT Run 5*	Baseline	0.365	Search by relevant terms and Human filtering

* These runs were submitted after the competition.

The temporally organised listing of relevant images is displayed in the lower part of the screen (the result-display panel). Each relevant image is listed with an overview metadata as a form of context. This metadata is configurable to display various sources of information, as required. Figure 2 shows a basic form of such metadata.

4 Exploiting LIFER for ImageCLEFlifelog2018 tasks

As mentioned, we exploited LIFER for ImageCLEFlifelog2018 LMRT and ADLT tasks. Firstly, based on the topic description, the search criteria are determined (by both automatically considering all the words in the queried topic as concepts, or alternatively allowing concepts to be ‘determined by the user). Secondly, we improved the interface of LIFER to allow user manually select multiple relevant images for the submission or taking all of them as relevant.

In term of ranking, we determine to use the default option of LIFER by chronological order.

In the next section, we present the official results on the test set exploiting the LIFER system.

Table 3. Selected criteria for the test set in ADLT task.

Topic	Activities	Times	Locations	Concepts
T001	-Airplane, -transport	+MinuteID: 400-1400	+Office	+vegetable, +Salad, +food
T002	-Walking, -Running, -Transport	+MinuteID: 720-1080(workday)	+Work, -Home	+Equipment
T003	-Running, -Transport	+MinuteID: 540-1080	-Restaurants, -Airport	+people, +indoor
T004	-Running, -transport	+MinuteID: 400-660(weekend)	+Work, +Home	+camera
T005	-Airplane, -Transport	+MinuteID: 540-1140(workday)	-Work, +Home	+dinner, +food
T006	-Airplane, -Transport	+MinuteID: 400-540(workday), 1140-1400(workday), 400-1400(weekend)	+Work, +Home	+furniture, +Chair, +Bed, +Cabinet
T007 u1	+Transport	+MinuteID: 400-540, 660-840, 1080-1190	Any	+indoor
T008	-Transport, -Running	+MinuteID: 400-1400	+Costa Coffee	+People
T009	+Airplane, +Transport	+MinuteID: 590-1400(workday), 540-1240(weekend)	Any	+cellphone, phone
T010	-Transport, -Airplane	+MinuteID: 960-1190	-Work, -Home	+Gravestone

+ means selection and – means exception.

5 Results

We submitted 10 runs (5 for each task) in total, summarized in Table 1 and 2. For ADLT, the best result is made by searching the relevant keywords and using human-in-loop to filter the irrelevant results. The remaining runs are created by using the relevant query terms and human filtering. For LMRT, the first two runs are automatic while the rest three adopted the same approaches used in ADLT. Shown in Tables 3 and 4 are the search criteria for the best run of each task.

6 Discussions and Conclusions

In this paper we introduce different baseline approaches, from fully automatic to fully manual approaches, by exploiting LIFER, an interactive lifelog retrieval

Table 4. Selected criteria for the test set in LMRT task.

Topic	User	Activities	Times	Locations	Concepts
T001	u1	-Running, -Transport	+MinuteID: 540-1140(workday)	+Work, -Home	+Coffee, +Cup
T002	u1	-Running, -Transport	+MinuteID: 400-540(15/08/2016 to 15/09/2016)	-Work, -Home	+Shop, +Store
T003	u1	-Transport, -Airplane	+MinuteID: 400-1400	-Work, +Home	+food, +vegetables, + Kitchen
T004	u1	-Transport, -Airplane	+MinuteID: 400-540(01/09/2016 to 30/09/2016)	+Home	+TV
T005	u2	-Waking, -Running, -Transport	+MinuteID: 400-1400(01/09/2016 to 30/09/2016)	+Work, -Home	+People, +Indoor
T006	u2	-Waking, -Running, -Transport	+MinuteID: 720-1400(weekend), 960-1400(workday)	Any	+Cellphone, Phone
T007	u2	-Transport, +Walking, -Running	+MinuteID: 400-1140(05/09/2016 to 09/09/2016)	+Work	-Computer, -Laptop + Office
T008	u2	+Walking	+MinuteID: 430-590, 1080-1190	Any	+ Street
T009	u3	-Transport, -Airplane	+MinuteID: 400-540, 660-840, 1080-1190	-Home, -Work	+church
T010	u3	-Transport,-Airplane	+MinuteID: 590 - 1140	-Home	+Restaurant

+ means selection and – means exception.

system to tackle the ImageCLEFlifelog 2018 task, as a participant of the Lifelog Moment Retrieval and Activities of Daily Living Understanding tasks. These approaches, that require different levels of involvement from the users, exploit only the information provided by the organizers along with the collection of images, e.g., the description of the semantic locations and the physical activities. With the human in the loop, we obtained the highest score for ADLT task (ADLT Run 1, *please notice that we are not ranked since we are the task organiser). However, without the manually input, the results can be close to random (as in the result of LMRT Run 1). This shows that the key challenge is how to translate the query to the search criteria, with requires further study.

References

1. Bolanos, M., Mestre, R., Talavera, E., Nieto, X.G., Radeva, P.: Senseseer mobile-cloud-based lifelogging framework. Visual summary of egocentric photostreams by representative keyframes pp. 1–6 (July 2015)
2. Byrne, D., Lavelle, B., Doherty, A.R., Jones, G.J., Smeaton, A.F.: Using bluetooth and gps metadata to measure event similarity in sensecam images. 5th International Conference on Intelligent Multimedia and Ambient Intelligence (July 2007)
3. Dang-Nguyen, D.T., Piras, L., Giacinto, G., Boato, G., De Natale, F.G.: Multimodal retrieval with diversification and relevance feedback for tourist attraction images. ACM Transactions on Multimedia Computing, Communications, and Applications (2017), accepted
4. Dang-Nguyen, D.T., Piras, L., Riegler, M., Zhou, L., Lux, M., Gurrin, C.: Overview of ImageCLEFlifelog 2018: Daily Living Understanding and Lifelog Moment Retrieval. In: CLEF2018 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Avignon, France (September 10-14 2018)
5. Dang-Nguyen, D.T., Riegler, M., Zhou, L., Gurrin, C.: Challenges and opportunities within personal life archives. In: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval. pp. 335–343. ICMR '18, ACM, New York, NY, USA (2018), http://doi.acm.org/10.1145/3206025.3206040
6. Doherty, A.R., Smeaton, A.F.: Automatically segmenting lifelog data into events. 9th International Workshop on Image Analysis for Multimedia Interactive Services (30 June 2008), http://doras.dcu.ie/4651/
7. Fan, C., Zhang, Z., Crandall, D.J.: Deepdiary: Lifelogging image captioning and summarization. Journal of Visual Communication and Image Representation (2018)
8. Fang, H., Gupta, S., Iandola, F., Srivastava, R., Deng, L., Dollar, P., Gao, J., He, X., Mitchell, M., Platt, J., Zitnick, L., Zweig, G.: From captions to visual concepts and back. IEEE Institute of Electrical and Electronics Engineers (June 2015), https://www.microsoft.com/en-us/research/publication/from-captions-to-visual-concepts-and-back/
9. Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., Gupta, R., Albatal, R., Dang-Nguyen, D.T.: Overview of NTCIR-13 Lifelog-2 Task. In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies (2017)
10. Gurrin, C., Schoeffmann, K., Joho, H., Dang-Nguyen, D.T., Riegler, M., Piras, L.: Lsc '18: Proceedings of the 2018 acm workshop on the lifelog search challenge. ACM, New York, NY, USA (2018)
11. Gurrin, C., Smeaton, A.F., Byrne, D., O'Hare, N., Jones, G.J.F., O'Connor, N.: An examination of a large visual lifelog. In: Li, H., Liu, T., Ma, W.Y., Sakai, T., Wong, K.F., Zhou, G. (eds.) Information Retrieval Technology. pp. 537–542. Springer Berlin Heidelberg, Berlin, Heidelberg (2008)
12. Gurrin, C., Smeaton, A.F., Doherty, A.R.: Lifelogging: Personal big data. Foundations and Trends in Information Retrieval 8(1), 1–125 (2014), http://dx.doi.org/10.1561/1500000033
13. Hong, M., Jung, J.J.: Visualizing multidimensional lifelogging data: A case study on mymoviehistory project. Cybernetics and Systems pp. 1–15 (2018)
14. Huang, C., Xu, H., Xie, L., Zhu, J., Xu, C., Tang, Y.: Large-scale semantic web image retrieval using bimodal deep learning techniques. Information Sciences 430, 331–348 (2018)

15. Ionescu, B., Lupu, M., Rohm, M., Gînsca, A.L., Müller, H.: Datasets column: diversity and credibility for social images and image retrieval. ACM SIGMultimedia Records 9(3), 7 (2018)
16. Ionescu, B., Müller, H., Villegas, M., de Herrera, A.G.S., Eickhoff, C., Andrearczyk, V., Cid, Y.D., Liauchuk, V., Kovalev, V., Hasan, S.A., Ling, Y., Farri, O., Liu, J., Lungren, M., Dang-Nguyen, D.T., Piras, L., Riegler, M., Zhou, L., Lux, M., Gurrin, C.: Overview of ImageCLEF 2018: Challenges, datasets and evaluation. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018), LNCS Lecture Notes in Computer Science, Springer, Avignon, France (September 10-14 2018)
17. Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.Y.: Multimodal deep learning. In: Proceedings of the 28th international conference on machine learning (ICML-11). pp. 689–696 (2011)
18. Peitgen, H., Jürgens, H., Saupe, D.: Chaos and fractals - new frontiers of science (2. ed.). Springer (2004)
19. Saritha, R.R., Paul, V., Kumar, P.G.: Content based image retrieval using deep learning process. Cluster Computing pp. 1–14 (2018)
20. Wang, P., Sun, L., Smeaton, A.F., Gurrin, C., Yang, S.: Computer vision for lifelogging: Characterizing everyday activities based on visual semantics. In: Computer Vision for Assistive Healthcare, pp. 249–282. Elsevier (2018)
21. Zhou, L., Dang-Nguyen, D.T., Gurrin, C.: A baseline search engine for personal life archives. In: Proceedings of the 2nd Workshop on Lifelogging Tools and Applications. pp. 21–24. LTA '17, ACM, New York, NY, USA (2017), http://doi.acm.org/10.1145/3133202.3133206
22. Zhou, L., Hinbarji, Z., Dang-Nguyen, D.T., Gurrin, C.: Lifer: An interactive lifelog retrieval system. In: LSC@ICMR (2018)