LifeSeeker 3.0 : An Interactive Lifelog Search Engine for LSC'21

Thao-Nhu Nguyen*
Tu-Khiem Le*
Van-Tu Ninh*
Dublin City University
Ireland

Minh-Triet Tran
Nguyen Thanh Binh
University of Science
Vietnam National University Ho Chi Minh city
Vietnam

Graham Healy
Annalina Caputo
Cathal Gurrin
Dublin City University
Ireland

ABSTRACT

In this paper, we present the interactive lifelog retrieval engine developed for the LSC'21 comparative benchmarking challenge. The LifeSeeker 3.0 interactive lifelog retrieval engine is an enhanced version of our previous system participating in LSC'20 - LifeSeeker 2.0. The system is developed by both Dublin City University and the Ho Chi Minh City University of Science. The implementation of LifeSeeker 3.0 focuses on searching and filtering by text query using a weighted Bag-of-Words model with visual concept augmentation and three weighted vocabularies. The visual similarity search is improved using a bag of local convolutional features; while improving the previous version's performance, enhancing query processing time, result displaying, and browsing support.

CCS CONCEPTS

• **Information systems → Multimedia databases**; **Users and interactive retrieval**; *Search interfaces*; • **Human-centered computing → Interactive systems and tools**.

KEYWORDS

lifelog, interactive retrieval, information system

ACM Reference Format:

Thao-Nhu Nguyen, Tu-Khiem Le, Van-Tu Ninh, Minh-Triet Tran, Nguyen Thanh Binh, Graham Healy, Annalina Caputo, and Cathal Gurrin. 2021. LifeSeeker 3.0 : An Interactive Lifelog Search Engine for LSC'21. In *Proceedings of the 4th Annual Lifelog Search Challenge (LSC '21), August 21, 2021, Taipei, Taiwan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3463948.3469065

1 INTRODUCTION

With the rapid development of sensor technology, commercial wearable devices are being widely used to capture many aspects of one's life, and serve many purposes, such as health analysis [34], sport performance analysis [2], or self-quantifying physical and mental health prediction [30]. Due to the growing popularity of using sensors and wearable devices for self-quantifying/self-tracking, passive image capturing and video recording, the vision of "The Memex Concept" originating by Bush in 1945 [4], which aims to create "an archive of a personal lifetime experience and knowledge storage for later usage and sharing", is being gradually implemented. Lifelogging, which began to develop since the capability of implementing "The Memex Concept" had become possible, has not been an active research topic until the seminal MyLifeBits [8] lifelog database in 2006 and the release of the upgraded MyLifeBits multifaceted retrieval system in the same year [8].

In recent years, many research challenges have been organized to introduce new problems in the lifelog domain to the research community; ImageCLEF [5–7, 29], NTCIR [9–11], and the Lifelog Search Challenge (LSC) [13]. Among the newly introduced problems, the Lifelog Moment Retrieval Task (LMRT) is an important one that has been attracting much attention from the lifelogging research community, with its vision to develop a lifelog-based memory prosthetic, as well as providing a deep insight into an individual's habits and activities. Among many lifelogging research challenges, the Lifelog Search Challenge (LSC) is the most competitive one, which aims to become an international competitive benchmarking activity to support the fair and accurate comparison of different approaches to interactive retrieval from lifelog datasets [14].

In this paper, we analyze LifeSeeker 2.0's performance during last year's challenge and introduce the LifeSeeker 3.0 - an enhanced version of this retrieval system. The new features developed in LifeSeeker 3.0 are as follows:

- **Metadata Enhancement:** We refine the semantic location from the provided metadata and manually cluster them into 32 categories. The city and country labels of the semantic locations are also added based on the given GPS data. In addition, more visual concepts are extracted automatically using the ResNet-101 model trained on the Visual Genome dataset [20] and Microsoft Vision API.
- **Image Visual Similarity Matching:** We construct a Bag-of-Features representation using Local Convolutional Features (LCF) for visual similarity search, which has been competitive on the Oxford and Paris buildings benchmark [26].
- **Weighted Bag-of-Words for Free-text Search:** We weight the terms in the L2-norm Bag-of-Words query vector during the cosine similarity computation to re-rank the retrieval results for the free-text search features. The weights (w) are based on the importance of terms in three predefined vocabularies: time, location (loc), and visual concepts (vc) (with $w_{time} > w_{loc} > w_{vc}$). Thereby, the multifaceted filter panel is also attached to the free-text search features and

*Three authors contributed equally to this research.

LSC '21, August 21, 2021, Taipei, Taiwan

ACM ISBN 978-1-4503-8533-6/21/08…$15.00
https://doi.org/10.1145/3463948.3469065

time/location information can be used to retrieve the desired moments precisely.

2 RELATED RESEARCH

Image retrieval [17] has seen a significant increase in research interest over the past decade. At the early stages, retrieval techniques were based on textual features, which first annotate images with text and then search by using text-based methodologies [16, 27]. Since manual annotation of large modern datasets is time-consuming and labor-intensive, the emergence of other worthwhile approaches which utilize the visual properties of mages, i.e. content-based image retrieval (CBIR) [18, 28, 33], became necessary. Especially, CBIR extracts visual concepts such as objects, colors, texts, and so forth in order to create metadata represented by a set of multi-dimensional vectors. The system converts different forms of user queries into vectors in the multi-dimensional space, where similarities are computed regarding the searching process. We rely extensively on such visual analytics.

The Lifelog Search Challenge (LSC), an annual image retrieval competition in the ACM International Conference on Multimedia Retrieval (ICMR), has been organized since 2018. In order to compete in the challenge, each team is required to build a real-time interactive tool that can complete both known-item and ad-hoc tasks over a large multimodal lifelog dataset. Among 14 teams partaking last year, we highlight three teams whose systems performed the best. For the very first time, Mysceal [32], the most effective system presented at LSC'20, introduced an efficient engine through the combination of Elastic Search and automatic query expansion. Besides developing a TFIDF-based similarity search procedure, Tran et al. exploited the use of the map as not only a filter but also a movement visualization [32]. This engine outperformed all other competitors as the one with the lowest searching time and the highest number of occurrences of top-3 correct submissions. With very close results, SOMHunter [25], an adoptive video retrieval system developed for the Video Browser Showdown (VBS) [31] re-purposed for the LSC'20 task, ranked second place. SOMHunter enabled users to flexibly query by presenting the users with three different visualizations of the search results: (1) original ranked results, (2) re-scoring by user's relevance feedback, and (3) ranking using temporal features. Kovalčík et al. [19] proposed the VBS-inherited-system named VIRET [19], which was among the top-performing architectures also in the LSC'20 competition. Their query mechanism is considerably updated; users can search either traditionally by a textual query, visually by drawing a color, or semantically by sketching.

Like most traditional retrieval systems, LifeSeeker 3.0 is a search engine based on textual and visual features. The most remarkable contribution of the LifeSeeker 3.0 this year is to improve the free-text query by using a weighted Bag-of-Words model with visual concept augmentation and three different vocabularies in terms of time, location, and visual concepts. Besides, the image visual similarity matching also plays a vital part in the system implementation performance.

3 LIFELOG DATA FOR THE EXPERIMENT

The LSC'21 organizers used a slightly reduced version of the dataset employed for LSC'20 and generated a new set of information needs The dataset is a multimodal dataset that combines data of a single lifelogger from the three previous LMRT challenges, NTICR 2015, 2016, and 2018 [14]. It contains 114-day multimodal lifelog data captured and synchronized from both smartphone and multiple sensors recording continuously for several years. The lifelog images are fully anonymized to prevent any personal information leaking, which could be used to trace back the identity of both the lifelogger and other people appearing in the images. The metadata provided by the organizers is enriched to provide descriptive and temporal information for each moment. This includes the automatic extraction of location attributes, location categories, and objects in the images using computer-vision neural networks including Places365CNN [36] and Mask-RCNN [15, 35] pre-trained on the COCO dataset [23] respectively.

4 LIFESEEKER 2.0 IN LSC'20

4.1 Overview of LifeSeeker 2.0

We developed LifeSeeker 2.0 [22] for LSC'20, which was an enhanced version of its first implementation in 2019 [21]. The system was designed to maintain an innovative user interface and optimize the search time by minimizing redundant interactions (e.g. animations, selection boxes, filtering sliders, menu bar) while maximizing the amount of information displayed to the users. LifeSeeker 2.0 introduced a number of enhancements over the first version. Firstly, image thumbnail size was reduced to accommodate more images on screen, which facilitated a more efficient relevance judgement process by enabling the user to quickly skim through the retrieval results without having to scroll too much. Secondly, additional metadata was made available for each image (e.g. date, time, imagery details that match query description), which were displayed in a pop-up window as illustrated in Figure 1. The image's details and its temporal linkages and visual similarity relationships with other images were efficiently encapsulated within the pop-up window. Finally, the temporal linkage was one of our main contributions to LifeSeeker 2.0, which we named Elastic Sequencing, and which displayed the moments before and after the current image respectively. This allows users to adjust how far in time should the moment be shown. This was accomplished by setting a delta time (time difference between images) to adjust the sequence of temporal images to be displayed. The Elastic Sequencing allows for a quicker check of chronological events, which is particularly useful for certain queries that involve events and times.

Along with the improvements in the user interface, the core indexing and retrieving system of LifeSeeker 2.0 was also upgraded. In addition to the automatically generated visual concepts provided by the organizers (as described in Section 3), we further tagged the images with the concepts from the Visual Genome dataset [20] using the Bottom-Up Attention model [1], and extracted any text visible in the images using CRAFT [3]. Moreover, the GPS readings were synthesized into human-readable addresses to be used as part of the indexing process. For retrieval, the LifeSeeker 2.0 relied

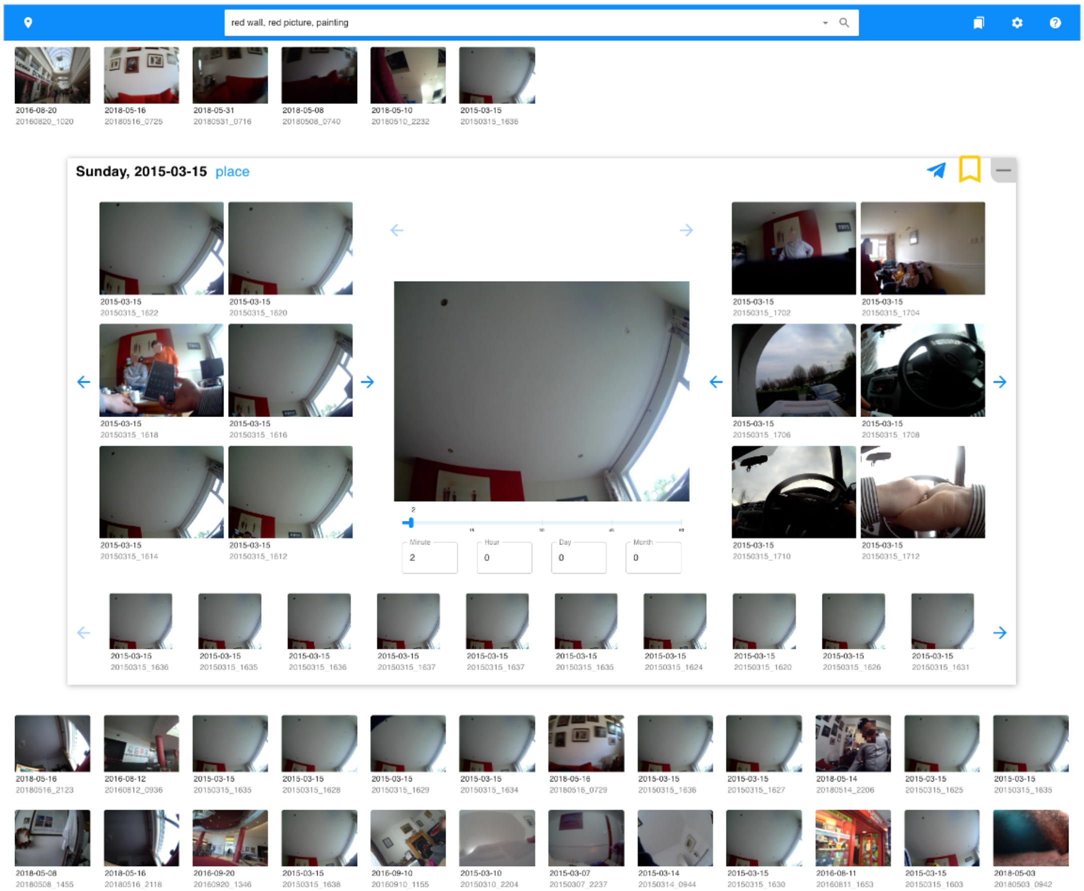

Figure 1: The interface of the system when searching for the query *"Four red figures, maybe they are aliens. It looked like a painting of aliens. There were walking on the desert. There was a big red wall behind the painting."*.
The details window shows the image's preview in the center, the left and right sections display the temporal linkage controlled by the Elastic Sequencing, the bottom are images that are visually similar to the current image.

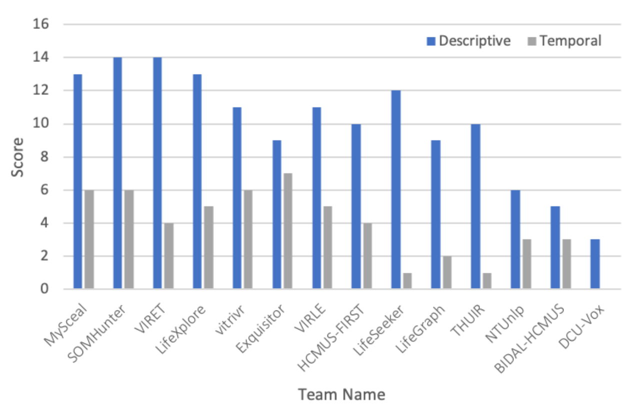

Figure 2: Performance comparison of all participating systems in terms of Descriptive topics and Temporal topics. The bars illustrates the number of correct submissions.

mainly on two search techniques: the Elastic Search and Bag-of-Words models. The users were able to use these interchangeably during their search.

In LSC'21, we continue improving this system by introducing a third version of LifeSeeker, which inherits the core functionalities of the 2.0 version, with minor fixes discussed in Section 4.2. LifeSeeker 3.0 is further enhanced with additional metadata and improved retrieval algorithms.

4.2 System Performance Analysis from LSC'20 Results

In the LSC'20 challenge, 24 queries were employed to evaluate 14 participating retrieval systems. Of these, 17 could be classified as descriptive topics and 7 classified as temporal topics. The descriptive topics describe the scenes and moments in great detail with visual clues; while the temporal topics emphasize the time sequences of activities. Figure 2, shows a summarised performance comparison of participating systems in LSC'20 by displaying the number of correct submissions of all teams, categorized by query type. LifeSeeker 2.0

achieved the overall score of 1,053 at the challenge and came mid-range among 14 participating systems ($\mu \simeq 1,038$, $\sigma \simeq 398$). The winning system achieved the highest score of 1,635. The overall score is calculated based on the search time and the number of incorrect submissions. A detailed description of how the score is calculated could be found in the LSC'18 overview paper [12]. The system was able to solve only 13 out of 24 topics (12 descriptive and 1 temporal queries, which resulted in a precision of 0.54). With respect to the top-5 performing systems in the challenge (Table 1), LifeSeeker was competitive in terms of descriptive topics, but performed poorly on the temporal ones.

Table 1: Performance comparison with top-performing teams in LSC'20 challenge

Team	All Topics	Descriptive	Temporal
MySceal	0.79	0.76	0.85
SOM Hunter	0.83	0.82	0.85
VIRET	0.75	0.82	0.57
lifeXplore	0.75	0.76	0.71
vitrivr	0.70	0.64	0.85
LifeSeeker (Ours)	0.54	0.70	0.14

From Table 1, it is clear that the precision of our system is negatively impacted by temporal topics. The significant difference between the scores for descriptive and temporal topics was caused by an error in the source code that affected the accuracy of temporal sequencing presented to the user. This issue also slightly influences LifeSeeker's performance on descriptive topics, as most of these include a temporal clue in the last expansion of the query. For example, the query: *"I see Steve Wozniac on a wall of portraits. The wall was a brick wall with a door and large heater. I was speaking to an audience before seeing the photos. I left by driving back to work. It was in 2015 in March on a Wednesday."*, contains an indication of time (Wednesday in March 2015), which helps to narrow down the search space and to identify the correct answer. Failing to deal with the temporal descriptions made LifeSeeker unable to retrieve the desired images.

5 OVERVIEW OF LIFESEEKER 3.0

In general, the user interface and interactive mechanism of LifeSeeker 3.0 are not extensively different from version 2.0, except that more details of a selected moment are shown. As illustrated in Figure 3, LifeSeeker 3.0 consists of three main parts: the indexed database, the interactive user interface of the search engine (client), and the server running retrieval algorithms which provides Application Programming Interfaces (APIs) for communications between the client and the server.

The indexed database consists of three weighted vocabularies (time, location, and visual-concepts); an inverted index; a listing of the top-50 visually similar search results of each lifelog image; and the pre-processed metadata in json format. The interactive user interface is kept the same as its previous version as illustrated in Figure 1; it has a free-text search bar with a list view of images showing the retrieval results. LifeSeeker 3.0 users have two options to search for the desired moment via server's APIs: Elastic Search or weighted Bag-of-Words; these can be toggled in the settings icon in the user interface. The weighted Bag-of-Words utilizes a conventional inverted index over the three pre-processed vocabularies extracted from the provided metadata. Elastic Search only requires a single json-format file containing all the pre-processed metadata.

5.1 Metadata Processing and Enhancement

Gaining insights from data is a key factor in every retrieval system and for the purpose of LSC'21, we mainly concentrate on revising the incorrectly located points, which consequently led to the identification of distant places having no connection with the correct answer. To be more precise, we refine the geographic coordinates of the given semantic locations and group them into 32 separate classes manually. Additionally, tagging cities and countries - especially for locations outside Ireland- could expedite the search process; while using elevation with a threshold could indicate whether the lifelogger is on a plane or not. On top of that, the Microsoft Vision API is used as an automatic feature extractor for lifelog images in order to expand the variety of visual concept annotations.

5.2 Visual Similarity Search using Local Convolutional Features

We derive this idea from the work of Mohedano et al., which uses the activation of the last convolutional layer of VGG16 network structure, pre-trained on ImageNet dataset, as the local features [26] for Bag-of-Features encoding instead of using SIFT features [24]. The dimension of the output of the activation function after going through the last convolutional layer is $N \times M \times D$. This means that it has D different $N \times M$ feature maps which can be considered as $N \times M$ local CNN features (descriptors) in dimension D. The procedure of using Bag-of-Features is the same as described in [22].

The visual similarity search employs the global search approach with global query expansion as described in [26]. In detail, the Bag-of-Visual-Words (BoVW) vector is built with all the local CNN features extracted from the queried image and is compared with other BoVW vectors using cosine distance to find the top seven most visually similar images in the corpus. These top-seven image BoVW vectors are then averaged with the original query image vector to generate a new vector representation. Finally, this vector is used to compute the similarity with all the images and generate the final list of the 50 top-ranked images.

5.3 Weighted Bag-of-Words for Free-text Search

Besides using Elastic Search for retrieval, we also implement a customized Bag-of-Words algorithm for both free-text search and filtering. In detail, we construct three vocabularies from the pre-processed metadata which include time, location, and visual concepts.

- **Time vocabulary**: contains the information of month (from January to December), weekday (from Monday to Sunday), part of the day (early morning, late morning, afternoon, etc.).
- **Location vocabulary**: contains the refined semantic location names, countries, cities, and place categories outputted from Places365CNN.

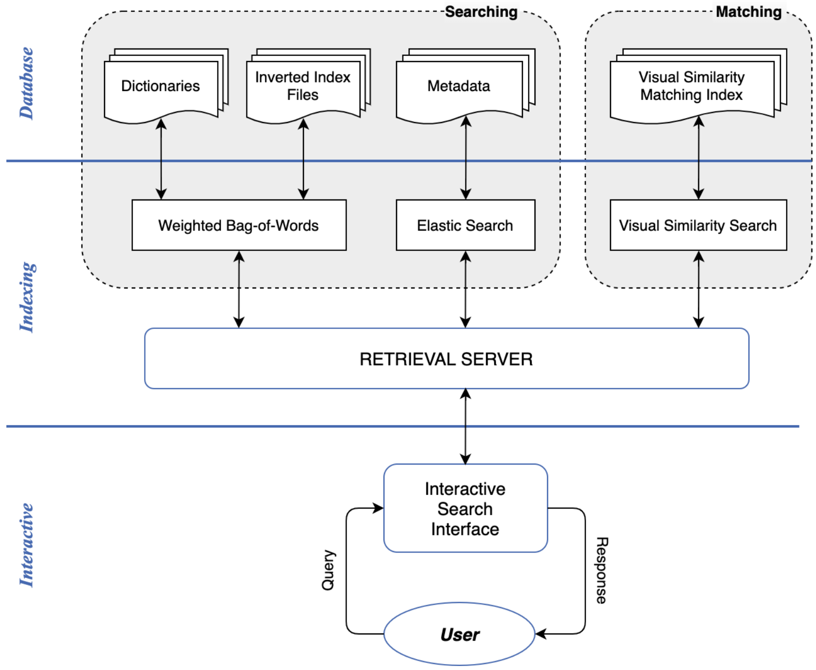

Figure 3: The system architecture and the workflow of LifeSeeker 3.0

- **Visual-concept vocabulary**: contains the object labels (MS-COCO and Visual Genomes dataset) detected automatically from the lifelog images using ResNet-101 model.

The vocabularies are refined so that there are no overlapping terms between the three vocabularies. In addition, we also manually filter the vocabularies to remove meaningless terms. The weight (w) for each vocabulary is also different. We select the weight based on the importance of the vocabularies. In this case, the weight of the terms in time vocabulary ($time$) is higher than the location one (loc). The location vocabulary is considered more important than the visual-concept one (vc), as it would be easier to navigate to the desired moment if the location is given in the query. These weights are combined into a vector, then it is multiplied into the L2-norm term frequency vector of the query. Thereby, the weights can amplify the time and location when computing cosine similarity between the query vector and the L2-norm term frequency vector of images in the archive to retrieve relevant images. In our algorithm, we do not consider inverse document frequency (IDF) weighting as the occurrences of terms in the corpus are all considered to be equally important. The IDF weighting would reduce the significance of some common terms which appear frequently in the lifelog annotation corpus such as week-of-day, part-of-day, and semantic location labels.

To summarize this idea, we use the following formula for our re-defined cosine similarity computation between the weighted Bag-of-Words query vector and the ones in the archive. Let $\boldsymbol{tf_q}$ and $\boldsymbol{tf_i}$ be the L2-norm term-frequency vector of the query and the L2-norm term-frequency vector extracted from the annotation of the lifelog image $\boldsymbol{i}$ in the dataset respectively. The term weighting of three vocabularies (time, location, and visual concepts) is represented by a vector $\boldsymbol{w}$. The score computed from our re-defined cosine similarity follows the formula:

$$\text{score} = \frac{(\boldsymbol{w} \odot \boldsymbol{tf_q}) \cdot \boldsymbol{tf_i}}{\|\boldsymbol{tf_q}\| \|\boldsymbol{tf_i}\|}$$

6 CONCLUSION

In this paper, we present - LifeSeeker 3.0 - a new interactive retrieval system for the LSC challenge that improves previous systems of the same name.

We briefly analyzed LifeSeeker 2.0's performance in LSC'20 based on the results provided by the organizers to identify potential opportunities for improvement.

While the user interface of LifeSeeker 3.0 has not changed, the underlying searching mechanism is improved with weighted Bags-of-Word for free-text search and visual similarity search using local convolutional features. With more concepts extracted from Microsoft Vision API, refined GPS data, and semantic location names, both the variety of visual-concept labels and the accuracy of filtering results using semantic location increases.

As a plan for future improvements, we are investigating approaches to build a semi-automatic label recommendation system that could propose unknown objects in lifelog images for annotation. Thereby, the lifelog images can have better tags, which will boost the search engine's performance. Moreover, we plan to conduct a user study to test the efficiency of our new searching algorithms in LifeSeeker 3.0 to identify further improvements that could be added to the system.

ACKNOWLEDGMENTS

This publication has emanated from research supported in party by research grants from Irish Research Council (IRC) under Grant Number GOIPG/2016/741, Science Foundation Ireland under grant numbers SFI/12/RC/2289_P2, SFI/13/RC/2106_P2 and 18/CRT/6223. This work was supported by the ADAPT Centre, Insight Centre for Data Analytics and Centre for Research Training in Artificial Intelligence funded by Science Foundation Ireland Research Centres Programme (Grant 13/RC/2106; 13/RC/2106_P2) and co-funded by the European Regional Development Fund.

REFERENCES

[1] Peter Anderson, Xiaodong He, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould, and Lei Zhang. 2018. Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering. In *CVPR*.

[2] Gobinath Aroganam, N. Manivannan, and David J. Harrison. 2019. Review on Wearable Technology Sensors Used in Consumer Sport Applications. *Sensors (Basel, Switzerland)* 19 (2019).

[3] Youngmin Baek, Bado Lee, Dongyoon Han, Sangdoo Yun, and Hwalsuk Lee. 2019. Character Region Awareness for Text Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 9365–9374.

[4] Vannevar Bush. 1945. As We May Think. *Atlantic Monthly* 176, 1, 641–649. https://doi.org/10.1145/227181.227186

[5] Duc-Tien Dang-Nguyen, Luca Piras, Michael Riegler, Giulia Boato, Liting Zhou, and Cathal Gurrin. 2017. Overview of ImageCLEFlifelog 2017: Lifelog Retrieval and Summarization. In *CLEF2017 Working Notes (CEUR Workshop Proceedings)*. CEUR-WS.org <http://ceur-ws.org>, Dublin, Ireland.

[6] Duc-Tien Dang-Nguyen, Luca Piras, Michael Riegler, Minh-Triet Tran, Liting Zhou, Mathias Lux, Tu-Khiem Le, Van-Tu Ninh, and Cathal Gurrin. 2019. Overview of ImageCLEFlifelog 2019: Solve my life puzzle and Lifelog Moment Retrieval. In *CLEF2019 Working Notes (CEUR Workshop Proceedings)*. CEUR-WS.org <http://ceur-ws.org>, Lugano, Switzerland.

[7] Duc-Tien Dang-Nguyen, Luca Piras, Michael Riegler, Liting Zhou, Mathias Lux, and Cathal Gurrin. 2018. 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.

[8] Roger Gemmell, Jim; Bell, Gordon; Lueder. 2006. My lifebits: a personal database for everything. *Commun. ACM* 49, 1 (2006), 88–95. https://doi.org/10.1145/1107458.1107460

[9] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatal. 2016. Overview of NTCIR-12 Lifelog Task. (2016), 354–360. http://eprints.gla.ac.uk/131460/ The authors acknowledge the financial support of Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289 and the input of the DCU ethics committee and the risk & compliance officer. We acknowledge financial support by the European Science Foundation via its Research Network Programme ?Evaluating Information Access Systems?.

[10] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Rashmi Gupta, Rami Albatal, and Duc-Tien Dang-Nguyen. 2017. Overview of NTCIR-13 Lifelog-2 Task. In *Proceedings of the 14th NTCIR conference (NTCIR-14)* (Tokyo, Japan).

[11] Cathal Gurrin, H. Joho, Frank Hopfgartner, Liting Zhou, Tu Ninh, Tu-Khiem Le, Rami Albatal, D.-T Dang-Nguyen, and Graham Healy. 2019. Overview of the NTCIR-14 Lifelog-3 task.

[12] C. Gurrin, Klaus Schoe, Mann, Hideo Joho, A. Leibetseder, L. Zhou, A. Duane, Duc-Tien Dang-Nguyen, M. Riegler, Luca Piras, M. Tran, Jakub Loko, and Wolfgang Hürst. 2019. Paper Comparing Approaches to Interactive Lifelog Search at the Lifelog Search Challenge (LSC 2018).

[13] Cathal Gurrin, Klaus Schoeffmann, Hideo Joho, Bernd Munzer, Rami Albatal, Frank Hopfgartner, Liting Zhou, and Duc-Tien Dang-Nguyen. 2019. A Test Collection for Interactive Lifelog Retrieval. In *MultiMedia Modeling*, Ioannis Kompatsiaris, Benoit Huet, Vasileios Mezaris, Cathal Gurrin, Wen-Huang Cheng, and Stefanos Vrochidis (Eds.). Springer International Publishing, 312–324.

[14] Cathal Gurrin, Klaus Schoeffmann, Bjorn Thor Jonsson, Duc Tien Dang Nguyen, Jakub Lokoc, Luca Rossetto, Minh-Triet Tran, Wolfgang Hurst, and Graham Healy. 2021. An Introduction to the Fourth Annual Lifelog Search Challenge, LSC'21. In *ICMR '21, The 2021 International Conference on Multimedia Retrieval*. ACM, Taipei, Taiwan.

[15] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. 2017. Mask R-CNN. *CoRR* abs/1703.06870 (2017). arXiv:1703.06870 http://arxiv.org/abs/1703.06870

[16] Eva Hörster, Rainer Lienhart, and Malcolm Slaney. 2007. Image Retrieval on Large-Scale Image Databases. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval* (Amsterdam, The Netherlands) *(CIVR '07)*. Association for Computing Machinery, New York, NY, USA, 17–24. https://doi.org/10.1145/1282280.1282283

[17] K. Juneja, A. Verma, S. Goel, and S. Goel. 2015. A Survey on Recent Image Indexing and Retrieval Techniques for Low-Level Feature Extraction in CBIR Systems. In *2015 IEEE International Conference on Computational Intelligence Communication Technology*. 67–72. https://doi.org/10.1109/CICT.2015.92

[18] K. Juneja, A. Verma, S. Goel, and S. Goel. 2015. A Survey on Recent Image Indexing and Retrieval Techniques for Low-Level Feature Extraction in CBIR Systems. In *2015 IEEE International Conference on Computational Intelligence Communication Technology*. 67–72. https://doi.org/10.1109/CICT.2015.92

[19] Gregor Kovalčík, Vít Škrhak, Tomáš Souček, and Jakub Lokoč. 2020. VIRET Tool with Advanced Visual Browsing and Feedback. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge* (Dublin, Ireland) *(LSC '20)*. Association for Computing Machinery, New York, NY, USA, 63–66. https://doi.org/10.1145/3379172.3391725

[20] Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A Shamma, Michael Bernstein, and Li Fei-Fei. 2016. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations. https://arxiv.org/abs/1602.07332

[21] Tu-Khiem Le, Van-Tu Ninh, Duc-Tien Dang-Nguyen, Minh-Triet Tran, Liting Zhou, Pablo Redondo, Sinead Smyth, and Cathal Gurrin. 2019. LifeSeeker: Interactive Lifelog Search Engine at LSC 2019. In *Proceedings of the ACM Workshop on Lifelog Search Challenge* (Ottawa ON, Canada) *(LSC '19)*. Association for Computing Machinery, New York, NY, USA, 37–40. https://doi.org/10.1145/3326460.3329162

[22] Tu-Khiem Le, Van-Tu Ninh, M. Tran, Thanh-An Nguyen, Hai-Dang Nguyen, Liting Zhou, G. Healy, and C. Gurrin. 2020. LifeSeeker 2.0: Interactive Lifelog Search Engine at LSC 2020. *Proceedings of the Third Annual Workshop on Lifelog Search Challenge* (2020).

[23] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. *CoRR* abs/1405.0312 (2014). arXiv:1405.0312 http://arxiv.org/abs/1405.0312

[24] David G. Lowe. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision* 60, 2 (Nov. 2004), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[25] František Mejzlík, Patrik Veselý, Miroslav Kratochvíl, Tomáš Souček, and Jakub Lokoč. 2020. SOMHunter for Lifelog Search. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge* (Dublin, Ireland) *(LSC '20)*. Association for Computing Machinery, New York, NY, USA, 73–75. https://doi.org/10.1145/3379172.3391727

[26] Eva Mohedano, A. Salvador, Kevin McGuinness, F. Marqués, N. O'Connor, and Xavier Giro i Nieto. 2016. Bags of Local Convolutional Features for Scalable Instance Search. *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval* (2016).

[27] Shailendrakumar M. Mukane, Sachin R. Gengaje, and Dattatraya S. Bormane. 2014. A novel scale and rotation invariant texture image retrieval method using fuzzy logic classifier. *Computers & Electrical Engineering* 40, 8 (2014), 154–162. https://doi.org/10.1016/j.compeleceng.2014.06.006

[28] Sudipta Mukhopadhyay, Jatindra Kumar Dash, and Rahul Das Gupta. 2013. Content-based texture image retrieval using fuzzy class membership. *Pattern Recognition Letters* 34, 6 (2013), 646–654. https://doi.org/10.1016/j.patrec.2013.01.001

[29] Van-Tu Ninh, Tu-Khiem Le, L. Zhou, Luca Piras, M. Riegler, P. Halvorsen, M. Lux, M. Tran, C. Gurrin, and Duc-Tien Dang-Nguyen. 2020. Overview of ImageCLEF Lifelog 2020: Lifelog Moment Retrieval and Sport Performance Lifelog. In *CLEF*.

[30] Bens Pardamean, Haryono Soeparno, Arif Budiarto, Bharuno Mahesworo, and J. Baurley. 2020. Quantified Self-Using Consumer Wearable Device: Predicting Physical and Mental Health. *Healthcare Informatics Research* 26 (2020), 83 – 92.

[31] Klaus Schoeffmann. 2019. Video Browser Showdown 2012-2019: A Review. In *2019 International Conference on Content-Based Multimedia Indexing (CBMI)*. 1–4. https://doi.org/10.1109/CBMI.2019.8877397

[32] Ly-Duyen Tran, Manh-Duy Nguyen, Nguyen Thanh Binh, Hyowon Lee, and Cathal Gurrin. 2020. Myscéal: An Experimental Interactive Lifelog Retrieval System for LSC'20. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge* (Dublin, Ireland) *(LSC '20)*. Association for Computing Machinery, New York, NY, USA, 23–28. https://doi.org/10.1145/3379172.3391719

[33] V. A. Wankhede and P. S. Mohod. 2015. Content-based image retrieval from videos using CBIR and ABIR algorithm. In *2015 Global Conference on Communication Technologies (GCCT)*. 767–771. https://doi.org/10.1109/GCCT.2015.7342767

[34] Jake Wu, Min & Luo. 2019. Wearable technology applications in healthcare: A literature review. *Online Journal of Nursing Informatics* 23, 3 (2019).

[35] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. 2019. Detectron2. https://github.com/facebookresearch/detectron2.

[36] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. 2017. Places: A 10 million Image Database for Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2017).