MEMORIA: A Memory Enhancement and MOment RetrIeval Application for LSC 2023

Ricardo Ribeiro
rfribeiro@ua.pt
IEETA, DETI, LASI
University of Aveiro
Aveiro, Portugal

Luísa Amaral
luisaamaral9@ua.pt
IEETA, DETI, LASI
University of Aveiro
Aveiro, Portugal

Wei Ye
weiye@ua.pt
IEETA, DETI, LASI
University of Aveiro
Aveiro, Portugal

Pedro Iglésias
pedrohpi@ua.pt
IEETA, DETI, LASI
University of Aveiro
Aveiro, Portugal

Alina Trifan
alina.trifan@ua.pt
IEETA, DETI, LASI
University of Aveiro
Aveiro, Portugal

Antonio J. R. Neves
an@ua.pt
IEETA, DETI, LASI
University of Aveiro
Aveiro, Portugal

ABSTRACT

The continuous collection and storage of personal data, denoted Lifelogging, has gained popularity in recent years as a means of monitoring and improving personal health. One important aspect of lifelogging is the collection and analysis of image data, which can provide valuable insights into an individual's lifestyle, dietary habits, and physical activity. The Lifelog Search Challenge provides a unique opportunity to explore the state-of-the-art in lifelogging research, particularly in the area of egocentric image retrieval and analysis. Researchers can propose their approaches and compete to solve lifelog retrieval challenges and evaluate the effectiveness of their systems on a rich multimodal dataset generated by an active lifelogger with 18 months of continuous capture of lifelogging data. This paper presents the second version of MEMORIA, a computational tool developed to participate in the Lifelog Search Challenge 2023. In this new version, the information retrieval is based on the use of natural language search with the possibility to filter the results based on keywords and time periods. The system applies image analysis algorithms to process visual lifelogs, from pre-processing algorithms to feature extraction methods, in order to enrich the annotation of the lifelogs. This new version explores the use of a graph database, more detailed image annotation, and event segmentation, in order to improve the performance and user interaction. Experimental results of the user interaction with our retrieval module are presented, confirming the effectiveness of the proposed approach and showing the most relevant functionalities of the system.

CCS CONCEPTS

• **Information systems** → **Search interfaces**; • **Human-centered computing** → *Interactive systems and tools.*

LSC '23, June 12–15, 2023, Thessaloniki, Greece

ACM ISBN 979-8-4007-0188-7/23/06.
https://doi.org/10.1145/3592573.3593099

KEYWORDS

lifelog, lifelogging, image processing, image annotation, data retrieval, object detection, Machine Learning, Information Systems

ACM Reference Format:
Ricardo Ribeiro, Luísa Amaral, Wei Ye, Pedro Iglésias, Alina Trifan, and Antonio J. R. Neves. 2023. MEMORIA: A Memory Enhancement and MOment RetrIeval Application for LSC 2023. In *6th Annual ACM Lifelog Search Challenge (LSC '23), June 12–15, 2023, Thessaloniki, Greece.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3592573.3593099

1 INTRODUCTION

The rapid development of technology over the past several years has lead to the emergence of devices such as wearable technology and smartphones that make it possible to collect data about our activities and behaviors. This particular data acquisition process is called lifelogging, and it can be performed actively or passively, resulting in a collection of digital records, which are named lifelogs. Due to the information that they hold, these records can then be applied in a variety of ways, allowing the extraction of insights about a person's health, their memories or even their behaviors and routines. However, due to the variety of devices that can be used to capture these records, they tend to be highly heterogeneous, since they can include images, audio clips, coordinates, biometrical information or even documents [19]. In addition, a passive process of capture also produces a large quantity of data, increasing the difficulty of analyzing it.

To effectively utilize the lifelog records, it is essential to organize, process, and retrieve the data efficiently, given the multimodal and large nature of the collection. These tasks are typically performed by a lifelogging system that extracts valuable information from each record, allowing for the search and retrieval of specific events in the lifelogger's digital memory archive.

The management and retrieval of collected lifelogs present a complex challenge that has garnered increasing attention in recent years. To promote research in this field and establish benchmark assessments for the developed systems, several lifelogging retrieval tasks have been introduced. One of those tasks is the Lifelog Search Challenge (LSC) workshop, which offers a live interactive search challenge where each competing team answers queries using the dataset of lifelogs provided by the organizers of the competition, and the team that responds with the greatest speed and accuracy wins the competition.

In this paper, we present the second prototype of MEMORIA (Memory Enhancement and MOment RetrIeval Application), which will be presented at the LSC'23 challenge [6]. After participating in the previous edition of the LSC challenge, LSC'22, we identified several performance and structural issues with the initial version of MEMORIA. In response, we made significant changes to various components of the system, such as pre-processing and image annotation. Additionally, we incorporated new features into the system, including event segmentation, a more advanced search engine, free-text search, and location processing. The changes include an expansion of the image annotation block to include optical character recognition, the YOLOv7 object detection model, the deep annotation models GRiT, and CLIP, a multimodal embedded model for generating captions.

A new event segmentation block employs a hierarchical method based on semantic locations and other image annotations. In free-text search, users can write their own queries, moving away from categorization filters. The search engine was also upgraded, switching from a relational database to a graph database and using full-text search. Furthermore, new semantic location annotations were added through GPS data clustering and a deep learning model for transport mode classification.

Performance improvements were also made to the MEMORIA user interface, such as the addition of thumbnails and a simpler upload system. With these improvements, MEMORIA is expected to achieve better retrieval performance for lifelog images from the LSC'23 dataset.

The paper is structured as follows. Following an introductory section, Section 2 presents recent work and challenges that have fostered research in lifelogging. Section 3 describes the MEMORIA changes and improvements implemented. Section 4 presents some results of user interaction with the system. Finally, Section 5 provides concluding remarks and outlines future work.

2 RELATED SYSTEMS

The most recent LSC edition [5] received the participation of nine distinct lifelog retrieval systems, which included MEMORIA, a newcomer in the competition. This edition saw a huge increase in the use of multimodal embedding approaches, with seven teams using OpenAI's Contrastive Language-Image Pre-Training model [16] to enhance the retrieval abilities of their systems.

Vitrivr [7] relies on a database management system with a focus on multimedia retrieval to carry out boolean and vector-space retrieval. This made it possible to retrieve information using boolean queries that incorporate metadata, as well as to retrieve vector textual embeddings that were extracted from the images using CLIP [16]. Vitrivr-VR [20] shares the same backend as the prior system, but it offers a more interactive interface that enables multimodal query formulation and result exploration by interacting with real-world objects within a virtual reality interface.

Memento 2.0 [2] enhanced its user interface and improved a functionality where it was possible to search for an event by using another temporally close event as context. This system also leveraged the embeddings provided by two distinct CLIP models to develop an ensemble ranking mechanism, where the score of similarity between an image and the query would be a combined rating of both the model's outputs.

Voxento 3.0 [3] offered a voice-based retrieval mechanism, improving the accessibility of lifelog retrieval, while sharing the same backend as Memento. The system also provided the option of employing a text-based search approach to complement this method. By utilizing textual information found in images and scene contexts that can distinguish between a home or work environment, Voxento also improved the automatic selection of filters to use throughout the search process, decreasing the number of results.

FIRST 3.0 [9] initially uses concepts to filter the images and then complements this approach by leveraging CLIP embeddings to explore the remaining results. It does this by extracting features from each image at various levels of detail, allowing the construction of a more detailed set of embeddings. The system is also augmented with an external search engine that allows searching for visual examples of concepts that are unfamiliar to it. Lastly, to facilitate data exploration, the images are grouped in hierarchical clusters based on their visual similarity and location.

Similarly, LifeSeeker 4.0 [15] also employed a CLIP model to extract a global meaning from each lifelog image, instead of concepts, and uses the extracted embeddings to cluster visually similar images in groups. In addition, with the introduction of lifelogs related to music in this edition of the competition, this system took advantage of this data to estimate the mood of the lifelogger during specific events.

By substituting an embedding strategy for a concept-based approach, E-myscéal [22] improved its search engine's ability to more accurately match each query's semantic content to the images. The system's authors ran an experiment using queries from the previous challenge edition to demonstrate how this change affected the retrieval process. They came to the conclusion that using these models significantly decreased the amount of hints required to retrieve the answers to the queries. Similar to other systems, visual concepts are still utilized to filter the results and save computational space. Two additional alternative visualization techniques that aid in the presentation of images that were visually similar or captured at the same time were added to the system, improving the user interface.

LifeXplore [10] was the other system, besides MEMORIA, that didn't use a multimodal embedding approach. The system relies on the extraction of visual concepts from images using the Myscéal team's Microsoft Cognitive Services outputs [23] and YOLOv4. Similar to MEMORIA, metadata like date, time, and location are also included to enhance the image annotations. To find visually related photos, Inception Net v3 [21] deep feature vectors are also used.

The performance of each system in the competition lead to the identification of the main features that a lifelogging system should possess to perform a fast and effective retrieval. Events segmentation can save screen space and optimize the search process, reducing the retrieval time, and the use of state-of-the-art CLIP models and embeddings approaches proved to be very useful. Lastly, an uncluttered and easy-to-use interface is also a key component in facilitating the search of lifelogs.

3 MEMORIA OVERVIEW

After participating in the previous edition of the LSC workshop [5], the MEMORIA system underwent major improvements to enhance its performance and usability. This section provides an overview of the changes made to MEMORIA, which include significant modifications to its architecture. Specifically, the database and search engine were modified, and the image annotation and pre-processing blocks were expanded. Free-text search was included, allowing users to create their own queries. The goal of these improvements was to enable the efficient organization, processing, and retrieval of the multimodal and large collection of lifelog data, ultimately enhancing the system's ability to retrieve relevant events from the LSC'23 dataset. The pre-processing block was kept to reduce the amount of images presented to the user based on a BIQA (Blind Image Quality Assessment) algorithm [17, 18].

3.1 Semantic Location Annotation

In addition to the lifelog images, the LSC'23 challenge dataset includes metadata about GPS locations, time, physiological data, among others. The previous version of MEMORIA did not take into consideration the semantic location annotations, which affected the system's performance. Since semantic locations are crucial for the retrieval task, the GPS coordinates were processed to enrich the annotations and to be used in the new event segmentation block.

MEMORIA is now capable of annotating semantic locations by filtering and clustering GPS coordinates for each day. A geocoding reverse API provides addresses for the resulting clusters. The data clustering method is performed using the HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) clustering algorithm [11]. Figure 1 displays an example of the clusters generated by MEMORIA based on GPS coordinates.

To extract further insights from the GPS data, we developed a deep learning model for transport mode classification, trained on the Geolife GPS trajectory dataset [26]. The model recognizes different modes of transportation, such as car, walk, bus, bicycle, among others. In addition, to associate the GPS data with the lifelog images, we developed a temporal synchronization mechanism, which allows us to annotate the images with the corresponding location and transport mode information. This synchronization enables efficient retrieval of the lifelog images based on location and transportation mode.

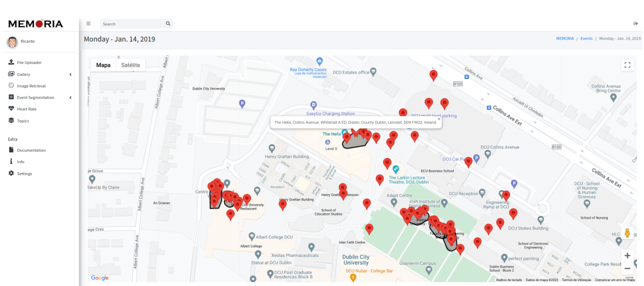


Figure 1: Example of clusters based on GPS coordinates annotated by MEMORIA.

3.2 Image annotation

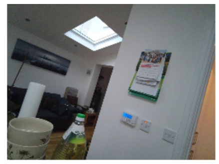
A crucial aspect of a lifelog retrieval system is the quality of the information that is extracted from each lifelog image, since it facilitates the process of searching for specific details or general events in the collection of digital records. This information should be adequate to the use case in mind, and it should be organized in the most effective way. The previous version of the MEMORIA system adopted a concept based approach to annotate the lifelog images. Multiple deep learning models were used to perform a range of computer vision tasks, such as determining the quality of an image, detecting the objects depicted in it and understanding the context of its scene.

The new version expands the tools used in this approach, and it also complements it with an embedded model, supporting a more detailed image annotation pipeline. Regarding the first approach, an optical character recognition model [4] was added to identify and detect characters in the lifelogs, complementing the OCR data that is already supplied by the LSC organizers. In addition, the YOLOv5 model [1] that was previously being used for object detection was substituted by the newer model YOLOv7 [24] to obtain more accurate detections. A deep annotation model [25] was also introduced in the system to further increase the level of detail of the annotations. This new model enables the detection of not only objects classes but also of their attributes, activities they are involved in and spatial relationships between them. These updates improved detection confidence, introduced the identification of words and characters, and provided more detailed annotations for activity, attribute, and spatial distribution extraction.

To complement the concept-based approach, the multimodal embedded model CLIP [16] was introduced in the annotation pipeline. Examples of these new annotations in images of the last edition's dataset can be observed in Figure 2, as well as the textual captions that were generated for these specific images.

The CLIP model proved itself to be very popular and successful in the previous edition of the competition due to its ability of learning the relationship between visual and semantic knowledge provided by the scene depicted in an image. However, while the other teams utilized the image embeddings generated by CLIP to compare them with the embeddings of the textual queries, MEMORIA 2.0 does not adopt this approach and instead leverages the fact that these image embeddings can be employed to generate a caption that describes each image, obtaining a more semantically rich understanding of the lifelog. Mokady et al. [14] proposes an approach that takes advantage of the semantic features of a VIT-B/32 CLIP model to provide textual context to an image by automatically generating a natural language text segment that describes it. This model was integrated into MEMORIA, allowing the generation of a caption for each uploaded image, which is then stored and used to enrich the annotations extracted from it. This approach allies itself to the enhancement that the new version of MEMORIA presents in the processing of free text, rather than just concepts and keywords, since the generation of captions requires an understanding of richer semantic textual content.

Types of annotations	MEMORIA at LSC'22	MEMORIA at LSC'23
Object detection	[(car, 0.638) , (car, 0.399)]	[(car, 0.621), (truck, 0.663)]
OCR	———————	[]
Caption generation	———————	"view from the driver's seat."
Deep annotation	———————	[("red truck in rear", 0.554), ("the rear view mirror on the car", 0.554)]



Types of annotations	MEMORIA at LSC'22	MEMORIA at LSC'23
Object detection	[(bottle, 0.904), (couch, 0.663), (potted plant,0.389), (refrigerator, 0.367)]	[(tv, 0.471), (couch, 0.906), (cup, 0.432), (cup, 0.353), (bottle, 0.922)]
OCR	———————	[('flora', 0.836),('10o', 0.387), ('gooune', 0.063)]
Caption generation	———————	"view from the kitchen into the living room."
Deep annotation	———————	[("a bottle of water", 0.607), ("a white bucket on the floor", 0.737), ("a calendar page with scenery on it", 0.6803), ("a skylight in the ceiling", 0.699), ("a white and blue label on a refrigerator", 0.615), ("an electrical outlet on the wall", 0.595), ("a newspaper on the wall", 0.620), ("green cap on a water bottle", 0.5839)]

Types of annotations	MEMORIA at LSC'22	MEMORIA at LSC'23
Object detection	[(person , 0.499), (car, 0.370)]	[(person , 0.43), (car , 0.46)]
OCR	———————	[("centra", 0.166)]
Caption generation	———————	"driving a car in the rain."
Deep annotation	———————	[("a man driving a car", 0.640), ("person in the car", 0.564), ("the hand of a person", 0.626)]

Types of annotations	MEMORIA at LSC'22	MEMORIA at LSC'23
Object detection	[(car, 0.938), (car, 0.914), (car, 0.313)]	[(car, 0.963), (car, 0.916), (car, 0.298)]
OCR	———————	[('o6di', 0.258), ('58377', 0.433)]
Caption generation	———————	"property image # detached holiday home in quiet location with garden"
Deep annotation	———————	[("a blue car parked on the street", 0.702), ("black car parked in street", 0.670), ("license plate on the car",0.707), ("red tail light on a car", 0.582), ("back tire of the car", 0.614), ("the back window of the car", 0.589), ("window on a brick building", 0.606), ("window on a building", 0.626), ("window on a building", 0.587), ("cars parked in front of a building", 0.609), ("a red brick building", 0.551), ("window of a blue van", 0.585), ("window in a brick building", 0.645), ("window on a building", 0.622)]

Figure 2: Improvements in the annotations captured from LSC'22 dataset images.

3.3 Segmentation of events

As lifelog data is usually passively captured and continuous, the temporal sequence of events is an important factor in any lifelogging system. To address this, MEMORIA's latest version includes an event segmentation method based on hierarchical events, leveraging data previously processed in the system, such as location, activity, and transport mode annotations related to lifelog images.

The first layer of event segmentation consists of daily events, covering all moments of a person's day, from waking up until going to sleep. In the second layer, each daily event is temporally segmented based on semantic locations and other annotations. Event segmentation provides information on the temporal sequencing of moments, enabling the system to retrieve and filter moments based on their temporal sequences, such as the moment before or after. An example of event segmentation for a part of a day can be seen in Fig. 3.

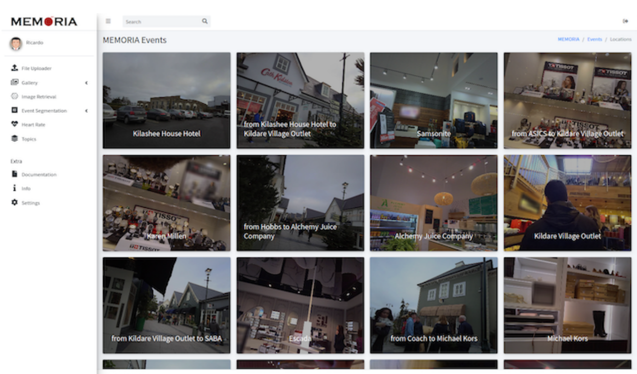


Figure 3: Example of the event segmentation for a part of one day, based on locations annotated by MEMORIA.

3.4 Free-Text Search

The previous version of MEMORIA used only categorization filters for searching. It divided the search into categories such as objects, activities, locations, environment, date, and part of the day, with a confidence threshold selector and a cumulative search that included the results of the previous search. Word expansion was also employed to increase the number of search results. In the new version, users have the option to use free text search, allowing them to write their own queries and explore the lifelog data in a more flexible way. The cumulative search feature from the previous version has been retained and extended with a history table where users can view, delete, and navigate between their past searches. The search function now passes the text query through the processing pipeline, which includes pre-processing, annotation, and indexing, to generate relevant search results.

The processing pipeline is divided into three phases: tokenization, expansion, and query. In the tokenization phase, raw text is divided into tokens, which include objects, events, activities, locations, dates, times, and temporal aspects of the sentence using keywords such as "after" and "before". The expansion phase expands the tokens found using word2vec [13]. In the query phase, the expanded tokens are used to create a database query that is passed on to the search engine to retrieve the desired results.

3.5 Search Engine

In this version of MEMORIA, we have explored the use of a different search engine. A comparative study between Neo4j and PostgreSQL was conducted by the authors in [12] for their multimedia retrieval system, which showed that Neo4j outperformed PostgreSQL. Inspired by these findings, we performed a similar test to compare the performance of both databases in MEMORIA.

We inserted around 440K images, 10K different tags, and additional data into both databases. In Neo4j, we ended up with approximately 33 million nodes and 140 million relationships, while PostgreSQL had around 40 million rows of data. We tested multi-table joins, few-table joins, and full-text search. Our findings show that Neo4j generally outperformed PostgreSQL in MEMORIA, particularly in tasks such as retrieving images associated with specific tags or uploaded by specific users and associated with specific tags generated by specific models.

The authors of [8] have proposed a method to significantly improve the speed of Lucene's full-text search using specific hardware. We were able to optimize our full-text search speed based on this method without requiring additional hardware, resulting in a 40% reduction in search time. Subsequently, we created a plugin with the optimized version of Lucene and integrated it into Neo4j.

In the previous version of MEMORIA, it was possible to retrieve images by specific tags, but the images could not be sorted by the number of related tags. However, it is desirable for images containing both tags to have a higher score and appear before those containing only one of them. This sorting can be effectively achieved through full-text search.

3.6 Performance

Besides what have been described in this section, in order to enhance the performance of the MEMORIA user interface, several blocks have been optimized. Firstly, thumbnails have been added to complement the original lifelog images. During the upload process, images are resized to a smaller resolution, which eliminates the need for the user interface to perform this operation when requested by the user.

Additionally, the previous upload system has been replaced with a more user-friendly one, allowing users to easily drag and drop folders of images and files into the system or select a system folder for MEMORIA to search for images to be uploaded.

4 RESULTS

In this section, an example of the MEMORIA user interaction is presented using an LSC'22 topic. In the LSC, a topic is presented to the participants at certain timestamps in seconds. Each task has a duration of 3 minutes (180 seconds) and so progressively it is shown more information about the topic being searched (every 30 seconds). The following topic, where t is the timestamp that the information is shown, was used as an example:

t=0: I think it was the second time I visited the house with the stone shed / hovel.
t=30: The shed was under green trees
t=60: I was at an outside shopping mall
t=90: and driving a long way
t=120: that morning
t=150: in June 2020.

In MEMORIA, the user has to introduce a sentence on the retrieval interface. This sentence is processed by extracting key components, that were previously referenced in the Subsection 3.4. The according images are then retrieved and shown. Additional searches can be performed to combine with previous searches as a single search. Each subsequent search will affect the results of the previous search. Post-search filters can be applied to the current search in order to reduce and specify the results if they are too diverse. These filters can include a tag, a location, a date or a month or a year, a specific part of a day (morning, afternoon, among others), a day of the week, and the minimum image quality of the search results can also be set (see an example in Figure 4).

At any time, images can be selected and individually zoomed in, providing information about the tags associated with the selected image, as illustrated in Figure 5. Additionally, the user can visualize

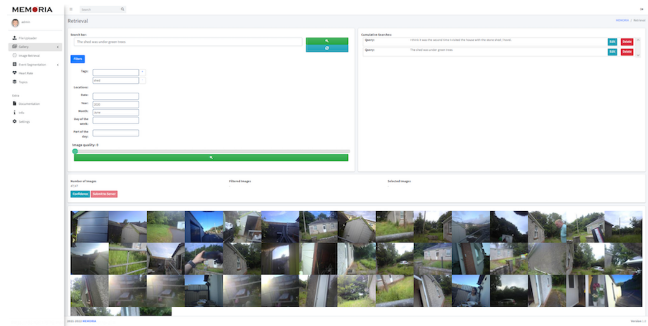


Figure 4: User interface of the retrieval view with two searches made with the following sentences "I think it was the second time I visited the house with the stone shed / hovel" and "The shed was under green trees" and the obtained search results.

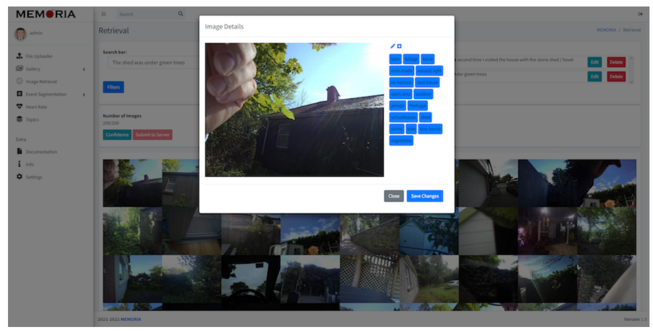


Figure 5: Zoomed-in image selected from the results of two searches made with the following sentences, "I think it was the second time I visited the house with the stone shed / hovel" and "The shed was under green trees".

images that were taken before or after the selected image, as shown in Figure 6. This is important because it allows the user to consider images in their temporal order and their appearance over time.

5 CONCLUSION AND FUTURE WORK

In this paper, the second prototype of MEMORIA (Memory Enhancement and MOment RetrIeval Application) was presented, with the intent to participate at LSC'23. Compared to the previous version, several changes have been made in the MEMORIA system to improve the results obtained in the previous LSC'22 challenge.

The main goal of the participation in LSC is to evaluate the improvements of the system and receive feedback for future improvements, learn from this research community and gain knowledge.

In this new version, some future work described in previous version have been addressed, such as implementing new computer vision algorithms to extract visual concepts from images to enrich the annotations, developing an event segmentation method based on information extracted from the lifelogs, adding a free-text search to the interface and making use of information like GPS coordinates that can be uploaded to the MEMORIA system.

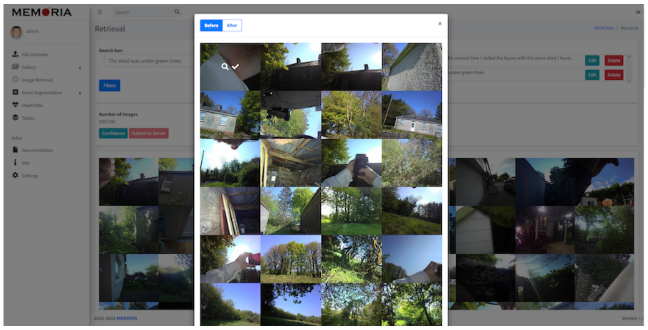


Figure 6: Image visualization of "one hour before" feature of a selected image from the results of two searches made with the following sentences, "I think it was the second time I visited the house with the stone shed / hovel" and "The shed was under green trees".

For future work, with the feedback and evaluation that will be provided in the LSC'23 competition, more improvements are intended, with the update of the system with new algorithms and mechanisms.

ACKNOWLEDGMENTS

This work was supported in part by National Funds through the FCT - Foundation for Science and Technology, in the context of the project UIDB/00127/2020, and by the project "Odyssey: Platform for Automated Sensing in Archaeology" (Ref. ALG-01-0247-FEDER-070150), co-financed by COMPETE 2020 and Regional Operational Program Lisboa 2020, through Portugal 2020 and FEDER.

REFERENCES

[1] 2022. ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation. https://doi.org/10.5281/ZENODO.7347926
[2] Naushad Alam, Yvette Graham, and Cathal Gurrin. 2022. Memento 2.0: An Improved Lifelog Search Engine for LSC'22. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 2–7. https://doi.org/10.1145/3512729.3533006
[3] Ahmed Alateeq, Mark Roantree, and Cathal Gurrin. 2022. Voxento 3.0: A Prototype Voice-Controlled Interactive Search Engine for Lifelog. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 43–47. https://doi.org/10.1145/3512729.3533009
[4] Youngmin Baek, Bado Lee, Dongyoon Han, Sangdoo Yun, and Hwalsuk Lee. 2019. Character Region Awareness for Text Detection. https://doi.org/10.48550/ARXIV.1904.01941
[5] Cathal Gurrin, Liting Zhou, Graham Healy, Björn Þór Jónsson, Duc-Tien Dang-Nguyen, Jakub Lokoć, Minh-Triet Tran, Wolfgang Hürst, Luca Rossetto, and Klaus Schöffmann. 2022. Introduction to the Fifth Annual Lifelog Search Challenge, LSC'22. In *Proceedings of the 2022 International Conference on Multimedia Retrieval* (Newark, NJ, USA) *(ICMR '22)*. Association for Computing Machinery, New York, NY, USA, 685–687. https://doi.org/10.1145/3512527.3531439
[6] Cathal Gurrin, Björn Þór Jónsson, Klaus Schöffmann, Duc-Tien Dang-Nguyen, Jakub Lokoč, Minh-Triet Tran, Wolfgang Hürst, Luca Rossetto, and Graham Healy. 2023. Introduction to the Sixth Annual Lifelog Search Challenge, LSC'23. In *Proc. International Conference on Multimedia Retrieval (ICMR'23)*. ACM, Thessaloniki, Greece.
[7] Silvan Heller, Luca Rossetto, Loris Sauter, and Heiko Schuldt. 2022. Vitrivr at the Lifelog Search Challenge 2022. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 27–31. https://doi.org/10.1145/3512729.3533003
[8] Jun Heo, Jaeyeon Won, Yejin Lee, Shivam Bharuka, Jaeyoung Jang, Tae Jun Ham, and Jae W. Lee. 2020. IIU. In *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems*. ACM. https://doi.org/10.1145/3373376.3378521
[9] Nhat Hoang-Xuan, Hoang-Phuc Trang-Trung, E-Ro Nguyen, Thanh-Cong Le, Mai-Khiem Tran, Tu-Khiem Le, Van-Tu Ninh, Cathal Gurrin, and Minh-Triet Tran. 2022. Flexible Interactive Retrieval SysTem 3.0 for Visual Lifelog Exploration at LSC 2022. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 20–26. https://doi.org/10.1145/3512729.3533013
[10] Andreas Leibetseder, Daniela Stefanics, and Klaus Schoeffmann. 2022. LifeXplore at the Lifelog Search Challenge 2022. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 48–52. https://doi.org/10.1145/3512729.3533005
[11] Leland McInnes, John Healy, and Steve Astels. 2017. hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software* 2, 11 (2017), 205.
[12] Nikolaj Mertz, Björn Þór Jónsson, and Aaron Duane. 2022. NeoCube. In *Proceedings of the 2nd International Workshop on Interactive Multimedia Retrieval*. ACM. https://doi.org/10.1145/3552467.3554799
[13] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. https://doi.org/10.48550/ARXIV.1301.3781
[14] Ron Mokady, Amir Hertz, and Amit H. Bermano. 2021. ClipCap: CLIP Prefix for Image Captioning. https://doi.org/10.48550/ARXIV.2111.09734
[15] Thao-Nhu Nguyen, Tu-Khiem Le, Van-Tu Ninh, Minh-Triet Tran, Thanh Binh Nguyen, Graham Healy, Sinéad Smyth, Annalina Caputo, and Cathal Gurrin. 2022. LifeSeeker 4.0: An Interactive Lifelog Search Engine for LSC'22. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 14–19. https://doi.org/10.1145/3512729.3533014
[16] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. https://doi.org/10.48550/ARXIV.2103.00020
[17] Ricardo Ribeiro, Alina Trifan, and António JR Neves. 2022. Impact of Blind Image Quality Assessment on the Retrieval of Lifelog Images. In *Proceedings of the 2nd International Workshop on Interactive Multimedia Retrieval*. 25–31.
[18] Ricardo Ribeiro, Alina Trifan, and António JR Neves. 2023. Blind Image Quality Assessment with Deep Learning: A Replicability Study and Its Reproducibility in Lifelogging. *Applied Sciences* 13, 1 (2023), 59.
[19] Ricardo Ribeiro, Alina Trifan, António JR Neves, et al. 2022. Lifelog Retrieval From Daily Digital Data: Narrative Review. *JMIR mHealth and uHealth* 10, 5 (2022), e30517.
[20] Florian Spiess and Heiko Schuldt. 2022. Multimodal Interactive Lifelog Retrieval with Vitrivr-VR. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 38–42. https://doi.org/10.1145/3512729.3533008
[21] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the Inception Architecture for Computer Vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2818–2826. https://doi.org/10.1109/CVPR.2016.308
[22] Ly-Duyen Tran, Manh-Duy Nguyen, Binh Nguyen, Hyowon Lee, Liting Zhou, and Cathal Gurrin. 2022. E-Myscéal: Embedding-Based Interactive Lifelog Retrieval System for LSC'22. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 32–37. https://doi.org/10.1145/3512729.3533012
[23] Ly-Duyen Tran, Manh-Duy Nguyen, Nguyen Thanh Binh, Hyowon Lee, and Cathal Gurrin. 2021. Myscéal 2.0: A Revised Experimental Interactive Lifelog Retrieval System for LSC'21. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 11–16. https://doi.org/10.1145/3463948.3469064
[24] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. 2022. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. https://doi.org/10.48550/ARXIV.2207.02696
[25] Jialian Wu, Jianfeng Wang, Zhengyuan Yang, Zhe Gan, Zicheng Liu, Junsong Yuan, and Lijuan Wang. 2022. GRiT: A Generative Region-to-text Transformer for Object Understanding. *arXiv preprint arXiv:2212.00280* (2022).
[26] Yu Zheng, Hao Fu, Xing Xie, Wei-Ying Ma, and Quannan Li. 2011. Geolife GPS trajectory dataset - User Guide. *Geolife GPS trajectories* (July 2011).