Considering Manual Annotations in Dynamic Segmentation of Multimodal Lifelog Data

Rashmi Gupta
Insight Centre for Data Analytics
School of Computing, Dublin City University
Dublin, Ireland
rashmi.gupta3@mail.dcu.ie

Cathal Gurrin
Insight Centre for Data Analytics
School of Computing, Dublin City University
Dublin, Ireland
cathal.gurrin@dcu.ie

***Abstract*—Multimodal lifelog data consists of continual streams of multimodal sensor data about the life experience of an individual. In order to be effective, any lifelog retrieval system needs to segment continual lifelog data into manageable units. In this paper, we explore the effect of incorporating manual annotations into the lifelog event segmentation process, and we present a study into the effect of high-quality manual annotations on a query-time document segmentation process for lifelog data and evaluate the approach using an open and available test collection. We show that activity based manual annotations enhance the understanding of information retrieval and we highlight a number of potential topics of interest for the community.**

***Index Terms*—Lifelogging, Event Segmentation, Information Retrieval**

I. Introduction

Lifelogging is the process of gathering large volumes of continuous multi-sensor personal data about an individual (including sequential images from wearable cameras) by using one, or more sensing devices [1]. Here, we need to consider how to segment such large lifelogs into manageable discrete units. In lifelogging, the contiguous set of indexable documents that have typically been combined into a logical unit called an event, which is created in a process called event segmentation [2], which is related to the topic of event detection. Event detection of continuous data streams has been the subject of research for two decades, in areas such as photo or video retrieval, and in many application domains it is seen as a solved problem. However, after a decade of lifelog data analytics, the segmentation of lifelog data streams into discrete documents is still a challenge. It is our conjecture that this is due to many reasons such as a lack of available datasets for comparative evaluation, or a lack of high-quality metadata upon which segmentation algorithms can be built.

For lifelog data, there exists a gap (i.e. in terms of retrieving semantic meaning and descriptive metadata) between human-labeled metadata annotations [3] and the performance of automatic tools [2], [4]–[10]. Therefore, in this work we explore the hypotheses that low-quality automated visual content annotations are a limiting factor for the effective segmentation of multimodal lifelog data. We evaluate this hypothesis by comparing automated metadata generation approaches with human annotated metadata in an experiment to segment semantically meaningful document units from continuous streams of multimodal lifelog data. We show that better quality annotations significantly enhance the quality of document segmentation of lifelog data and we use this to motivate the need for additional metadata and more research effort into the automatic annotation of lifelog data.

This paper's contribution is an analysis of the impact of enhancing the quality of metadata when generating retrievable document units from continuous stream lifelog data. In order to achieve this, we present a novel query-time segmentation process for lifelog documents that dynamically generates length and relevance optimized ranked lists of 'events' that are returned in response to a user query. This query-time approach replaces the conventional indexing-time event segmentation approach, by focusing on generating event segments that exactly match a user's query. We compare three alternative approaches to this segmentation of lifelog data streams; segmentation based on state-of-the-art visual concepts and other metadata (i.e. automatic annotations); segmentation based on manual annotations of human activities; and segmentation based on the fusion of both manual and automatic data sources. These three approaches are evaluated using the publicly available LSC2018 lifelog dataset, with 24 new topics that span a range of broad to narrow focus.

II. Related Work

As stated, event segmentation has been a focus of research for over two decades. Zacks and Tversky in 2001, define the event as a segment of time at a given location that is conceived by an observer to have a beginning and an end [11]. In this work, event segmentation refers to the process whereby continuous stream of multimodal wearable sensor data such as images/videos from wearable cameras; physical movement and biometric data from activity trackers; and semantic locations from location loggers etc. is segmented into discrete document units. Early work on the topic used basic metadata, such as colour change within wearable camera images over time, or movement metadata, to automatically segment lifelog data into indexable units [3], yet the authors noted a high degree of human subjectivity variance in the segmentation process and further approaches were required. Byrne et al. [4], introduced an automatic segmentation technique based on five low-level

978-1-5386-9151-9/19/$31.00 ©2019 European Union

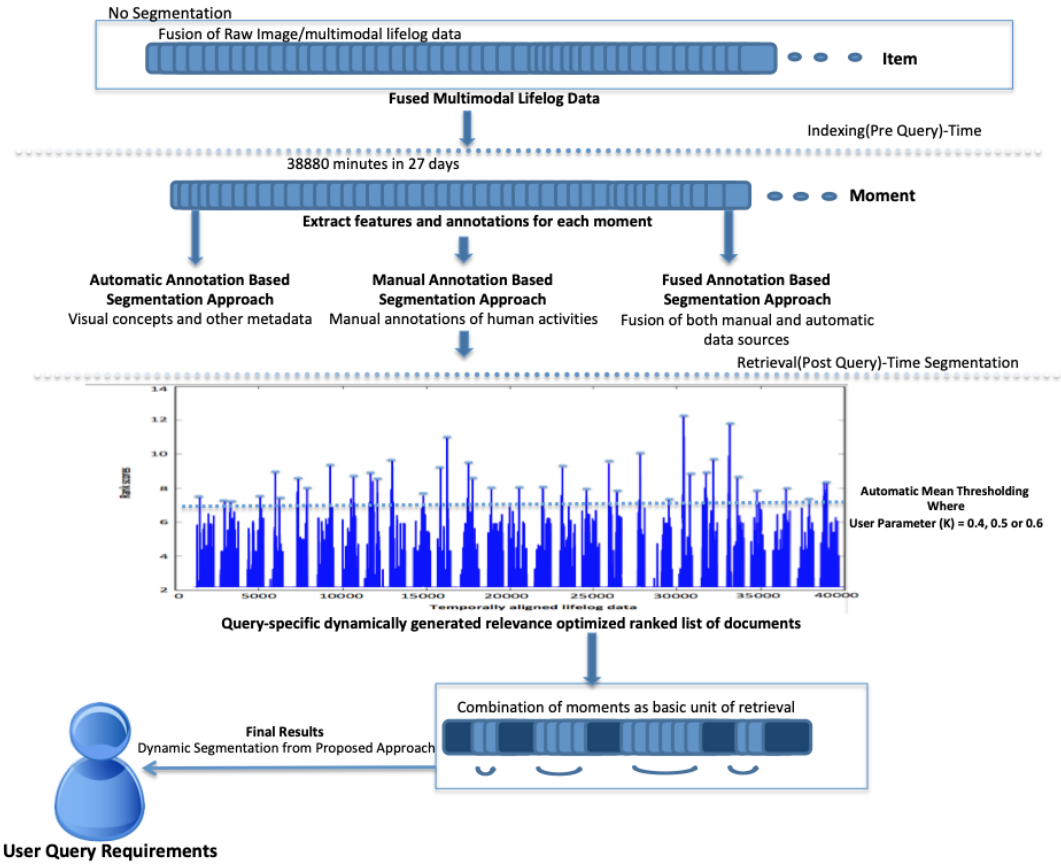

Fig. 1. Process of query-time multimodal lifelog segmentation for all three proposed approaches.

MPEG-7 visual descriptors from lifelog images along with contextual information such as change in light, human location or motion by using bluetooth and GPS metadata. Doherty et al. in 2008 [5], proposed a automatic event segmentation approach by using MPEG-7 visual descriptors, various vector distance methods and automatic thresholding techniques to enhance the performance of the segmentation. These initial approaches were characterized by the use of proprietary collections; while recently, reusable test collections have become the norm, such as CLEF [12], NTCIR [13], LSC [14] and EDUB-seg [15].

More recently, new automatic event segmentation approaches have emerged that utilize semantic visual concepts and location data [18], which report improved scores in the segmentation process. Molino et al. [19] proposed a new event segmentation approach which predicts upcoming next temporal segments based on the previous segments in continuous streams of lifelog data. In our previous work [2], we introduced two new segmentation approaches based on visual semantic concepts from Caffe framework [20] and high-quality image category labels for each lifelog image from the Microsoft Cognitive Services API [21]. We found that image category labels to be the best performing segmentation approach. In this work, we follow this trend and utilize automatic visual concepts as the source of data for our baseline segmentation process and compare this state-of-the-art approach to ones based on manual activity annotations. Unlike much of the prior work, our approach to event segmentation is novel in that it is a query-time process as opposed to the more conventional indexing time process. It is our belief that this leads to a more flexible segmentation algorithm and a comparison with the static segmentation approaches will be carried out at a later date. We point out that the contribution of this work is orthogonal to the choice of segmentation approach (indexing time or query time).

III. Query-time Segmentation Approaches

The basic premise of our work is based on our conjecture that a flexible query-time process should produce result documents that better fit a user's information need. Hence we have implemented a novel dynamic segmentation approach for this work. Our query-time segmentation process is (shown in Figure (1)) based on the following steps:

- Define a minimum document size (one minute in duration) called a moment and we fused the multimodal lifelog data into minute long segments (moments), which is our minimum indexable unit i.e. 1,440 moments/day and 38,880 moments in total for the LSC lifelog dataset that we use in this work [23].
- Extract automatic visual concepts from visual lifelog data (e.g. continuous stream wearable camera data) using a

modern concept detector. Additionally generate manual annotations of human activities for each moment in lifelog data (e.g. working, watching television, socializing, praying and/or travelling, etc.) where each moment may take place in a certain environmental settings/context (e.g. in an office environment, in a home, in a publicly-accessible building, etc.).
- Upon receiving a user query, generate a temporally-ordered query-similarity vector over all moments in the collection.
- Identify intra-moment similarity by implementing a standard distance method (e.g. Euclidean distance).
- Fuse similar scored sequential moments and declare a boundary if the similarity between moments is below a predefined threshold.

We built the search engine by implementing the Okapi BM25 ranking model to facilitate user queries. The model indexed moments as documents. Each moment was represented by a textual description extracted from the dataset metadata, that included date/time features, user activity logs, music listening history, biometric data, semantic locations, manual diet log and weather. This represented the basic lifelog metadata that we indexed in this experiment. This data was extended by one of three approaches (automatic, manual and fusion annotation) as described below. Full details of the metadata is discussed in the Dataset section below.

For the **automatic annotation** based segmentation approach, we appended visual concepts and descriptions generated by a high-level visual concept detector [21] (i.e. Automatic Annotation Based Segmentation Approach in Figure (1)) to the moments for indexing. This concept detector operates over every image in the lifelog data and produces high-quality (i.e. introducing new meaningful visual features) semantic metadata such as background color, foreground color, dominant color, description of particular image including caption, categories (i.e. based on 86-hierarchical categories taxonomy for each image) and tags. For example, the categorization of image is based on 86 super categories and further subdivided into detailed sub categories such as animal category (includes animal_bird, animal_cat, animal_horse or animal_panda), building category (includes building_arch, building_brickwall, building_stairs, building_church), transport category (includes trans_bicycle, trans_bus, trans_car and trans_trainstation); and the tags include objects (i.e. laptop, television, chair, table, knife), living beings (i.e. person, child, man, woman), scenery (i.e. sky, lawn, green, residential, sea) or actions (i.e. working, standing, sidewalk) that are relevant to the content of the particular image of lifelog data. An example image with high-quality visual concepts and descriptions is shown in Figure (2).

The **manual annotation** based segmentation approach amended annotations from an ontology of 24 real-world life activities that were labelled by a manual review of the moments by one expert researcher over a number of days. The ontology was a single-level ontology and included activity concepts (i.e. Manual Annotation Based Segmentation Approach in Figure (1)) such as: commuting to work, travelling, preparing meals, eating/drinking, taking care of children, praying, socializing/casual conversation, reading, gardening, shopping, work meetings, watching TV, playing computer games, using laptop/desktop computer, using mobile/tablet, any physical activity, sleeping, relaxing, organizing things, packing, cleaning, hygiene and make-up activity, writing on paper, searching/information seeking etc. An example image with manual annotations is shown in Figure (3).

Time: 11:19, Id: 1560451, DominantColorForeground: Grey, DominantColorBackground: Grey, accentColor: 893A61, isBWImg: false, DominantColors: [Grey], isAdultContent: false, isRacyContent: false, adultScore: 0.04483955, racyScore: 0.06651274, Description: {tags: [table, person, indoor, sitting, food, woman, coffee, plate, people, child, eating, laptop, meal, restaurant, group, man, dining, young, sandwich, girl, computer, holding, pizza, room, salad, phone], captions: [{ text: a group of people sitting at a table, confidence: 0.981832266 }] }, tags: [{ name: table, confidence: 0.9904625 }, { name: person, confidence: 0.97315377 }, { name: indoor, confidence: 0.958868146 }, { name: meal, confidence: 0.305454 }, { name: dining table, confidence: 0.114117384 }], categories: [{ name: others_, confidence: 0.01171875 }, { name: outdoor_, confidence: 0.00390625 }], faces: [{ age: 47.0, gender: true, faceRectangle: { left: 1951.0, top: 354.0, width: 280.0, height: 280.0 } }] }] }] }

Fig. 2. Examples of automatic high-quality visual concepts from concept detector using Microsoft Cognitive Service API [21].

Fig. 3. Example of manual annotations based on 24 real-world life activities.

For the **fusion annotation** approach, we utilized the fusion of both automatic and manual annotations and amended these to the metadata for indexing (i.e. Fused Annotation Based Segmentation Approach in Figure (1)) and provide a fair comparison with proposed segmentation approaches (discussed earlier in this section).

Once we had generated the moment textual annotation using one of the three approaches just described, a query-similarity vector was generated for all ranked and non-ranked moments and the Euclidean distance method was used to find the distance between each successive moment in the vector. We used an automatic mean thresholding technique (based on

mean, standard deviation and user parameter values (K = 0.4, 0.5 or 0.6)) to highlight the top ranked event boundaries (see formula below (1) and Figure (1)), based on the approach proposed in [5]. The value K was defined in an initial training phase that is not detailed here. In addition, we implemented an evaluation methodology discussed in [2], [5] that provides a fair and repeatable comparison among the new proposed approaches, which will be detailed in section IV.

$$Mean\_Threshold = mean + K * Standard\_deviation \quad (1)$$

A. Dataset Description

For this work we used the publicly available LSC2018 lifelog dataset [14] which consisted of 27 days of multimodal lifelog data (i.e. 38,880 moments in total) generated by one active lifelogger. Associated with the images (approx one per minute/moment) were various forms of metadata, such as date/time, high-level visual features (i.e. tags) for each image extracted from [21], and various other data sources that capture the real-world activities of the user. For this work we used music listening history such as song name, artist name, and album name; biometric data such as heart rate, galvanic skin response, sleep duration, calorie burn and steps count; semantic locations (e.g. home, work, restaurant); and a manual log of food and drinks. We also appended additional metadata to the collection that we know to be useful for retrieval, such as weather status (i.e. rain, fog, sunny, light showers) along with temperature conditions; day status (i.e. early morning, morning, afternoon, evening, night, late night), and the manual annotations that we mentioned previously. We are releasing these additional annotations (along with topics and relevance judgments described below) as an addendum to the LSC2018 collection in July 2019[1].

The example of identified automatic tags and manually annotated activities in visual lifelog data along with descriptive metadata is shown in Figure (4) below.

Fig. 4. Examples of multimodal LSC lifelog test collection along with manual annotations, automatic image tags and descriptive metadata.

[1]LSC2018 Dataset available at: http://lsc.dcu.ie. Additional annotations to be released in July 2019.

B. Queries

Although the LSC2018 collection includes 18 information needs, they were specifically designed to support interactive experimentation. Consequently we developed 24 new topics that reflect the wide range of query-types that would be expected to be used with lifelog collections, based on the proposals in [22], which motivates the development of different types of queries for lifelog retrieval systems. Consequently we developed 12 broad focus queries to support lifelog reminiscence/reflection, and 12 narrow-focus topics to reflect conventional retrieval needs (shown in Table (I)). Examples of broad queries include shopping, reading, working, driving, cleaning etc. Examples of narrow-focus topics include waiting for train, packing a suitcase, walking on a lovely day, eating an apple, brainstorming, having talk with a person who has ponytail. For each topic, we have manually generated complete relevance judgments which will also be released with the topics in the dataset addendum[1].

TABLE I
LIST OF 12 BROAD FOCUSED AND 12 NARROW FOCUSED USER QUERIES.

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

IV. EVALUATION METHODOLOGY

The main motive of this experiment was to evaluate the impact of annotation quality on the performance of a novel query-time event segmentation algorithm. In order to do this, we need to evaluate the quality of event boundary selection given the three different annotation methodologies described earlier. Since the event boundaries in lifelog data are inherently subjective, the evaluation methodology employed should be robust to minor variations in the boundary definitions. Therefore we employed a sliding window +/- 3 minutes that provides a necessary degree of flexibility in the measurement process. If a system defined event boundary is within 3 minutes of the human judgment, then it is considered to be accurate, which was the approach taken to segmentation evaluation in [5].

The top ranked moments were temporally clustered into candidate events and a mean thresholding method was employed to select only the highly ranked events for evaluation. The sliding window was used to identify how accurate the event boundary was and any boundary found that was greater

TABLE II
PERFORMANCE (I.E. PRECISION %) OF DYNAMIC EVENT SEGMENTATION BASED ON MANUAL ANNOTATIONS AND VISUAL CONCEPTS.

Broad Topic User Queries				Narrow Topic User Queries			
Topics	**Auto Annotation**	**Manual Annotation**	**Fusion**	**Topics**	**Auto Annotation**	**Manual Annotation**	**Fusion**
Shopping	0.81	0.99	0.99	Saturday morning coffee	0.99	1	1
Reading	0.90	0.99	0.99	Writing on paper	0.56	0.99	0.99
Cleaning	0.99	0.99	0.99	Waiting for a train	0.98	1	0.99
Socializing	0.98	0.99	0.99	Cutting the Grass	0.97	0.99	0.99
Driving	0.97	0.99	0.99	Packing	0.94	0.96	0.98
In a Meeting	0.96	0.99	0.98	Walking at Work	0.93	0.94	0.96
Watching TV at home	0.91	0.99	0.97	Walking to the Airplane	0.86	0. 99	0.88
Flying	0.81	0.96	0.96	DIY Store	0.78	0.93	0.87
Cooking at home	0.88	0.87	0.96	Eating an apple	0.67	1	0.86
Working	0.80	0.97	0.96	Fruit	0.67	0.76	0.72
Walking on a lovely day	0.75	0.86	0.89	Brainstorm	0.62	0.69	0.72
Resting	0.74	0.73	0.87	Ponytail	0.45	0.56	0.45
Average System Performance	**0.88**	0.94	**0.96**	**Average System Performance**	**0.86**	0.90	**0.88**

than +/- 3 minutes was considered to be incorrect. This methodology provides for precision values (see formula below (2)) to be calculated in terms of true positives, the negative effect of over segmentation (i.e. false positives) is ignored in this initial evaluation, as is the recall of lowly-ranked events which are relevant to the topic. The reason for this was a focus on finding (in most cases) the one and only relevant event for a given topic.

$$Precision(accuracy) = \frac{TruePositive}{TruePositive + FalsePositive} \quad (2)$$

V. RESULTS

We proposed and compared three different approaches as discussed earlier for all 12 broad and 12 narrow topic-specific queries (i.e. discussed in section III (B)), where we present and highlight the importance of manual annotations in automatic event segmentation approach to the research community.

- **Segmentation based on Visual Concepts:** A mean precision value 0.88 (for broad topics) and 0.86 (for narrow topics) in terms of overall accuracy of the segmentation approach based on the image tags description was calculated (summarized in Table (III)). In general, we found little difference between the broad and narrow topics when the retrieval was based on the automatic indexing process. There was a clear mismatch between the output (i.e. tags) of the concept detector and the topics as defined by the lifelogger. The narrow-focus topics such as eating an apple (0.67), brainstorming (0.62) and ponytail (0.45) proved to be comparatively difficult (see Table (II)). For example, the concept detector was labeling images as containing food, instead of apple or fruits; object based tags such as indoor, person, board etc. were semantically distant from the topics of the queries (e.g. work meeting), or the details of specific topics (e.g. talking to a man with a ponytail) were more detailed than the related concepts such as people/person. Ultimately the performance of the automatic segmentation approach could only be increased if, either there were more concepts, or the topics were somehow translated from the human semantic level into the system level, which is a variation of the well-known 'semantic gap' from multimedia retrieval.
- **Segmentation based on Manual Annotations:** Manual annotations, which naturally are more reflective of human semantic descriptions and consequently are more related to the human generated information needs. For topics with a clear semantic match with the human annotations, the result precision was significantly higher at 94%, which should not come as a surprise. However, where there was a descriptive difference between the scores for certain topics, such as cooking at home (0.87), walking on a lovely day (0.86), resting (0.73), looking at bowl full of fruits (0.76), brainstorming (0.69) and talking with a person who has ponytail (0.56) the precision levels were significantly reduced (see Table (II)). It is worth noting that these are mostly narrow-focus topics. The average precision of this approach is 0.94 for broad type user queries and 0.90 for narrow type user queries, which is clearly an improvement over the automatic annotation approach (summarized in Table (III)) and leads us to conclude that the accuracy of annotations is a significant factor in the overall performance of a segmentation algorithm.
- **Segmentation based on Annotations and Visual Concepts:** Combining both sources of metadata provided the best results with the score of precision 0.96 (increment of 0.08) for broad and slightly worse results 0.88 (increment of 0.02) for narrow type user queries (summarized in Table (III)). It is interesting to note that manual annotations improved the results for many of the poorly performing topics when compared to the automatic approach, such as writing on paper, eating an apple, shopping, flying in airplane, working on laptop and walking on lovely day (see Table (II)). This reinforces our hypothesis that enhancing the performance of the annotation engines has a significant impact of segmentation algorithm.

We found that consideration of manual annotations of human activities increases the precision scores of our approach to event segmentation approaches for both broad as well as narrow focus user queries. Although, we find a low increase (0.02) in narrow-focus user queries (discussed earlier), which

TABLE III
COMPARING SYSTEM PERFORMANCE (I.E. AVERAGE PRECISION) OF QUERY-SPECIFIC DYNAMIC EVENT SEGMENTATION BASED ON THREE DIFFERENT APPROACHES TO ANNOTATION

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

we can not claim to be significant. We can hypothesize that this is because the human annotations are both higher in quality, more accurate and more semantically meaningful than the automatic approaches. We also point out that this is an initial experiment and the positive effect of manual annotations is likely to decrease as more accurate and semantically meaningful automatic approaches to annotation of lifelog data become available in the coming years.

VI. CONCLUSION

In this paper we presented a study into the effect of incorporating manual annotations into the lifelog event segmentation process, which heretofore has been based on automatic annotations and automatically generated metadata. The study was carried out using a novel query-time segmentation model and evaluated using the reusable LSC2018 lifelog test collection with additional topics and metadata. Although we can suggest that all lifelogs can benefit from a human annotation of life activities, we realize that this is unlikely to occur in most real-world scenarios due to the human overhead of making manual annotations. Hence we propose that more detailed automatic annotations are required along with a research focus on developing activity-based annotations similar to the annotations generated by humans. For future work, we will explore the automation of this enhanced annotation process and we will also explore stricter evaluating criteria that penalize any missed boundaries.

VII. ACKNOWLEDGMENT

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289.

REFERENCES

[1] M. Dodge and R. Kitchin. outlines of a world coming into existence: Pervasive computing and the ethics of forgetting. Environment and Planning B: Planning and Design, 34(3):431445, 2007.11.

[2] R. Gupta, C. Gurrin, Approaches for event segmentation of visual lifelog data, in: MultiMedia Modeling. MMM 2018. Lecture Notes in Computer Science, vol 10704. Springer, Cham.

[3] A. R. Doherty, C. J. A. Moulin A. F. Smeaton (2011) Automatically assisting human memory: A SenseCam browser, Memory, 19:7, 785-795, DOI: 10.1080/09658211.2010.509732.

[4] D. Byrne, B. Lavelle, A. R. Doherty, G. J. Jones, and A. F. Smeaton. Using bluetooth and gps metadata to measure event similarity in sensecam images. 5th International Conference on Intelligent Multimedia and Ambient Intelligence, July 2007.9.

[5] A. R. Doherty and A. F. Smeaton. Automatically segmenting lifelog data into events. 9th International Workshop on Image Analysis for Multimedia Interactive Services, 30 June 2008.12.

[6] R. P. Bolaos M., Garolera M. Video segmentation of life-logging videos.In: PeralesF.J., Santos-Victor J. (eds) Articulated Motion and Deformable Objects. AMDO2014. Lecture Notes in Computer Science, pages 19, 2014.8.

[7] S. Karaman, J. Benois-Pineau, V. Dovgalecs, R. Mgret, J. Pinquier, R. AndrObrecht, Y. Gastel, and J.-F. Dartigues. Hierarchical hidden markov model in detecting activities of daily living in wearable videos for studies of dementia. Mul-timed Tools Appl(2014) 69:743.15.

[8] Y. J. Lee and K. Grauman. Predicting Important Objects for Egocentric Video Summarization. In: The International Journal of Computer Vision (IJCV), January 2015.

[9] A. G. d. Molino, B. Mandal, L. Li and L. J. Hwee. Organizing and retrieving episodic memories from first person view. In: 2015 IEEE International Conference on Multimedia Expo Workshops (ICMEW), June 2015.

[10] A. Furnari, S. Battiato and G. M. Farinell. Personal-Location-Based Temporal Segmentation of Egocentric Video for Lifelogging Applications. In: Journal of Visual Communication and Image Representation, 2018.

[11] M. J. Zacks, S. T. Braver, A. M. Sheridan, I. D. Donaldson, Z. A. Snyder, M. J.Ollinger, L. R. Buckner, and E. M. Raichle. Human brain activity time-locked to perceptual event boundaries. In: Nature neuroscience, pages 651655, 2001.19.

[12] H. Mller, P. Clough, T. Deselaers, B. Caputo, ImageCLEF: Experimental Evaluation in Visual Information Retrieval, Springer Publishing Company, Incorporated, 2010.

[13] C. Gurrin, H. Joho, F. Hopfgartner, L. Zhou, R. Gupta, R. Albatal, and D. T.Dang Nguyen. Overview of ntcir-13 lifelog-2 task. The Thirteenth NTCIR conference (NTCIR-13), dec 2017.13.

[14] D.-T. Dang Nguyen, K. Schoeffmann, and W. Hurst. Lsc2018 panel-challenges of lifelog search and access. In: ACM Workshop on The Lifelog Search Challenge,11-14 June 2018.10.

[15] Egocentric dataset of the university of barcelona segmentation (edub-seg) dataset, 2015. http://www.ub.edu/cvub/egocentric-dataset-of-the-university-of-barcelona-segmentation-edub-seg/.

[16] E.Talavera, M. Dimiccoli, M. Bolaos, M. Aghaei, P. Radeva. R-clustering for egocentric video segmentation. In Iberian Conference on Pattern Recognition and Image Analysis 2015 Jun 17 (pp. 327-336). Springer International Publishing.

[17] M. Dimiccoli, M. Bolaos, E. Talavera, M. Aghaei, S. G. Nikolov, P. Radeva. SR-Clustering: Semantic Regularized Clustering for Egocentric Photo Streams Segmentation. In Computer Vision and Image Understanding Journal (CVIU). 2015 Dec 22.

[18] S. Yamamoto, T. Nishimura, Y. Takimoto, T. Inoue and H Toda. PBG at the NTCIR-13 Lifelog-2 LAT, LSAT, and LEST Tasks, 2017.

[19] A. G. d. Molino, J.H. Lim and A.H. Tan. Predicting Visual Context for Unsupervised Event Segmentation in Continuous Photo-streams. In: ACM Multimedia Conference (MM '18), 2018.

[20] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadar-rama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093, 2014.14.

[21] H. Fang, S. Gupta, F. Iandola, R. Srivastava, L. Deng, P. Dollar, J. Gao, X. He,M. Mitchell, J. Platt, L. Zitnick, and G. Zweig. From captions to visual concepts and back. IEEE Institute of Electrical and Electronics Engineers, June 2015.

[22] S. Abigail and W. Steve. 2010. Beyond total capture: A constructive critique of lifelogging. Commun. ACM 53 (may 2010), 7077.

[23] R. Gupta, Considering documents in lifelog information retrieval, ICMR18. In: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval (2018) 497500.