

Activity Recognition and Segmentation Approaches to Multimodal Lifelog Data

Rashmi Gupta

Insight Centre for Data Analytics
Dublin City University
Dublin, Ireland
rashmi.gupta3@mail.dcu.ie

Cathal Gurrin

Insight Centre for Data Analytics
Dublin City University
Dublin, Ireland
cgurrin@computing.dcu.ie

Abstract—Lifelogging is a phenomenon whereby an individual digitally records his/her personal life experiences, for a variety of purposes. Activity recognition and segmentation is fundamental to many of the use cases in lifelogging. However, detecting sufficiently robust user activity boundaries that could be deployed with confidence in a subjective real-world setting remains a challenge. In this paper, we extend our previous work on identifying a better activity recognition and segmentation approach to multimodal lifelog data, primarily through the introduction of automatic thresholding techniques, but also through revising the criteria for selecting the most appropriate size of sliding window when evaluating the proposed algorithms. We use an open and publicly available lifelog test collection over a time period of 27 days with manual annotations and manually groundtruthed activities.

Index Terms—Lifelogging, Activity Recognition, Multimodal Lifelog Data Segmentation, Information Retrieval

I. INTRODUCTION

What if you will have a system on your computer that will allow you to retrieve all your past memories with your friends, family or colleagues? Imagine being able to tell every special life story or incident to your children or grandchildren. Lifelogging is a platform that proposes to make this possible. Already there are a number of different wearable devices available in the market to automatically capture our physical movements; text messages, emails or voice calls; web browsing; locations; sentimental situations; and our eating habits etc. So why not use all this personal archive to retrieve meaningful information? It is our belief that the activity recognition and segmentation is fundamental step to build such information retrieval engines in lifelogging. Thus in this paper, we used LSC2018 lifelog test collection [14] which includes wearable camera data, biometric data, descriptive/contextual metadata, and activity based manual annotations. In addition, we explored existing technologies to recognize and segment daily living user activities from continuous stream of multimodal lifelog data.

To access and retrieve meaningful information from multi-sensor lifelog data motivates the idea of developing better segmentation algorithms, in this paper. Prior research on lifelog data segmentation has taken an static (indexing-time) approach

which generates fixed inflexible indexable units, which are referred to as events. These events can vary in the range from a few minutes to a few hours. In our previous work, we inferred that the dynamic (query-time) segmentation approach would better support a user in searching for the specific content as per his/her information needs [16], [20]. We showed a number of retrieval scenarios where a predefined boundary (in case of static segmentation) is not optimal for the user, where such events are either too long or too short for the information needs.

Consequently, we implemented baseline static (indexing-time) segmentation approaches [11] that incorporate existing state of the art techniques [17]–[19] and newly proposed dynamic (query-time) segmentation approaches based on [16], [20]. We evaluate and compare both segmentation approaches using the publicly available LSC2018 lifelog dataset¹. We can identify our contributions thus: (i) identifying an optimal window size for the accurate detection of activity boundaries; (ii) refining our evaluation methodology for segmentation approaches to multimodal lifelog data; (iii) observing overall improvement of dynamic segmentation approach over static segmentation approach.

II. LITERATURE REVIEW

Activity boundary detection and segmentation is ubiquitous concept in lifelogging that is concerned with identifying regular (for example relaxing and watching television, eating food, working on laptop etc.) as well as occasional (for example attending birthday party or a cultural event etc.) life activities of the user. In this section, we explore all the existing state of the art segmentation approaches that motivate us to do the research on the study proposed in this paper.

Segmentation of lifelog data into manageable units of retrieval (for example events or activities) has been a focus of research for over two decades [17], [21], [22]. Here, an event is defined as a segment of time at a given location that is conceived by an observer to have a beginning and end [23]. An activity is defined as an un-interrupted sequential state of the individual in terms of their person or environment or stimuli [24]. Both represent a combination of sequential

minutes whose size is dependent on the activities of the individual, but an activity is considered to be shorter than an event, representing a physical activity, rather than a semantic action in a user's life. Based on past research, the event has been considered to be the main unit of retrieval from lifelogs and was employed manually in the initial SenseCam image viewer tool [1] as well as the early work in the development of lifelog search engines [2].

The automatic segmentation approach proposed in [3] helped the authors to note a high degree of human subjectivity variance in the definition of segmentation process and thus, further approaches were required. Byrne et al. [18], introduced an automatic segmentation approach based on MPEG-7 visual descriptors from lifelog image data along with contextual information such as change in light, human location or motion by using bluetooth and GPS metadata. Doherty et al. [17] proposed a similar approach based on MPEG-7 visual descriptors and enhanced the performance of segmentation algorithm by introducing automatic thresholding techniques and vector distance methods.

Yamamoto et al. [4], proposed an event segmentation approach based on features from visual concepts and location data when participating in NTCIR-13 lifelog-2 workshop [5]. Molino et al. [6], proposed a new indexing time segmentation approach to predict upcoming next temporal segments based on past segments and trained over the large archive of lifelog datasets (i.e. NTCIR [5], CLEF [7] and EDUB-seg [8]) and R3. Recently in 2018 [11], the authors have proposed two new indexing-time segmentation approaches using visual concepts from Caffe framework [9] and image categorization of the image from Microsoft cognitive services API [10]. They found the approach using image categorization from Microsoft cognitive service API as best performing approach for event segmentation task. The proposed approach is based on pre-existing static segmentation approach described in [17] and is reported to outperform the existing state of the art approaches [8], [12], [13], [17], [18].

In this work, we follow this trend and extract automatic visual concepts from Microsoft cognitive services API along with manual activity annotations as the source of data for both static (indexing-time) as well as dynamic (query-time) segmentation approaches. We show that dynamic segmentation approach leads to a more flexible (in terms of variable time duration of user activity as per his/her information needs) segmentation approach when compared with a state-of-the-art static segmentation approach. Thus, we consider that the contribution of this work is orthogonal to the choice of segmentation approach (indexing time or query time).

III. A METHODOLOGY FOR ACTIVITY RECOGNITION AND SEGMENTATION

The basic premise of our work is to identify a better segmentation algorithm that would accurately satisfy user information needs. Hence, we introduce a new methodology (shown in Fig. 1) for the proposed segmentation approaches (i.e. static and dynamic) that are based on following steps:

- 1) Collect multi-sensor lifelog data and define the smallest indexable unit of retrieval (one minute in duration) called a moment [24].
- 2) Fuse the multi-sensor lifelog data into minute long segments (moments). Consequently, we have 38,880 moments for 27 days in LSC2018 lifelog test collection.
- 3) Extract and generate the corpus of automatic visual concepts using one standard concept detector (for each image in lifelog dataset); manual annotations (for example eating food, relaxing, watching television at home, using laptop in an office environment etc.); and dataset metadata (for example date, time, age, height, weight, location etc.) for each moment (or image) in lifelog data.
- 4) Generate a query-similarity vector for all moments in the collection. In static segmentation, the vector is formalized with the normalized scores for all sequential activities performed in a day (for example min_id 300 to min_id 456 : having breakfast, min_id 457 to min_id 690 : commuting to work, min_id 691 to min_id 695 : walking at work, min_id 696 to min_id 900 : working on laptop at work, and so on for the remaining day) whereas in dynamic segmentation, the vector is formalized with the normalized scores for only one specific activity performed in a day (for example the driving a car activity has vector scores for min_id 457 to min_id 690 while commuting to work in the morning, and min_id 906 to min_id 1067 while driving home from work in the evening).
- 5) Implement a standard vector distance method (e.g. Euclidean distance) to identify the intra-moment similarity (for both static and dynamic approach).
- 6) Fuse similar scored sequential moments and declare a boundary if the similarity between moments is surpasses a predefined threshold.

We followed this proposed methodology in both static and dynamic segmentation approaches. In addition to this, we built an ad-hoc information retrieval engine by implementing the Okapi BM25 ranking model to facilitate user queries in dynamic segmentation approach. The segmentation model indexed all moments as lifelog documents [24] where each moment is represented by a textual description extracted from the dataset metadata. The dataset metadata include date and time features, semantic locations, biometric data, music listening history, user diet and mood log, and user activity based manual annotations. The detailed discussion about the data used in this study is explained (in dataset description section) below.

Once we had generated the textual description for each moment using metadata described above, a query-similarity vector is generated for all ranked (typically associated with normalized score values in 0 and 1) and non-ranked moments (usually zero scored values). A Euclidean vector distance method was implemented to identify the distance between each successive moment in the vector (see formula 1).

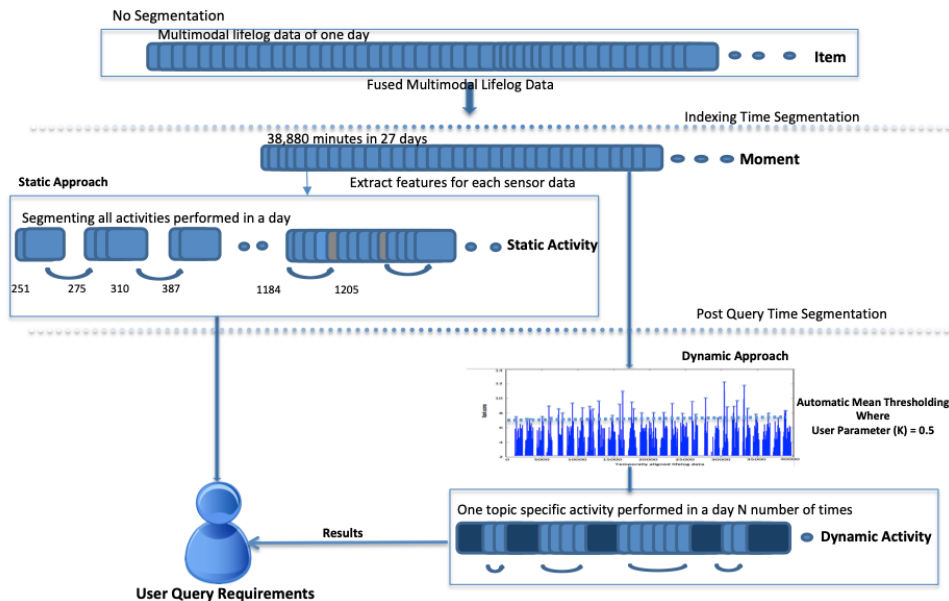


Fig. 1. Process of segmenting multimodal lifelog data in both indexing and query time segmentation approaches.

$$Euclidean_Distance(V) = \sqrt{\sum_{n=1}^{38880} (v_2 - v_1)^2} \quad (1)$$

We used automatic mean thresholding technique as (shown in the formula 2) to highlight the top ranked user activity boundaries (based on approach discussed in [11], [17]).

$$Mean_Threshold = mean + K * Standard_deviation \quad (2)$$

In addition, we propose an evaluation methodology based on the work in [11], [17] (a sliding window approach) to evaluate the approach performance. It also provides a fair and repeatable comparison among the new proposed approaches, which will be detailed in section IV. The main question here is what size should the sliding window be? Too small and the subjective human judgments as to semantic event boundaries will reduce the scored performance of an otherwise acceptable approach, too high and the scores could be artificially inflated due to larger window sizes.

IV. EVALUATION AND RESULTS

In this section, we will describe the dataset we used in this study and the query topics that reflect the wide range of information needs of the user. Also, we will discuss the evaluation method to compare static and dynamic segmentation approaches. On the behalf of that, we will observe which approach works better for the activity detection and segmentation task.

A. Dataset Description

In recent years, the two collaborative benchmarking exercises for comparative evaluation of approaches to various

lifelog access and organisation tasks were organised under the umbrella of NTCIR [5] and LSC [14], which provide dataset and queries that a lifelogger would wish to make against their collections. In this study, we are using publicly available LSC2018 lifelog test collection [14] which consisted of 27 days of multimodal lifelog data (typically 38,880 moments in total) generated by one active lifelogger.

The data used in LSC lifelog test collection was captured using a wearable camera, which usually captures about two images per minute from the point-of-view of the lifelogger. Associated with the images, were various forms of metadata such as date and time features associated with unique minute id; visual concepts (i.e. 86 taxonomy based categories) extracted for each image using [10]; and various other data sources that capture the real-world environmental/contextual information such as weather condition (for example rain, sunny, light showers etc.) and day status (for example early morning, morning, noon, evening, night, late night etc.). We embedded user's music listening history (for example song name, artist name, and album name); biometric data (for example heart rate, galvanic skin response, sleep duration, calorie burn and steps count); manual diet and mood log; semantic locations (for example home, work or restaurant). We also appended manual annotations in the feature vector (for each moment or minute or image) from an single-level ontology of 24 real-world user activities (for example 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.). These manual an-

notations were labelled by a trusted researcher. The example of all additional metadata sources used with LSC2018 lifelog dataset is shown in Fig. 2.

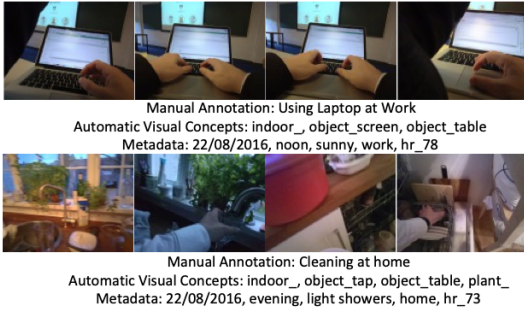


Fig. 2. Examples of multimodal LSC lifelog test collection along with manual annotations, automatic image categories and descriptive metadata.

B. Query Topics

Along with this dataset, we developed 24 new query topics, based on the proposals in [15]. Consequently, we developed 12 broad focus user query topics such as shopping, reading paper, cleaning, resting, driving car, travelling in airplane, working on laptop, socializing, in a meeting, preparing meals at home, watching television at home, walking on a lovely day etc. (developed to support lifelog reminiscence/reflection [15]) and 12 narrow focus user query topics such as Saturday morning coffee, walking to the airplane, waiting for a train at train station, cutting the grass, walking at work place, eating an apple, writing on paper, brainstorming, looking at fruit bowl, packing a suitcase, purchasing item in DIY store and having talk with person who has ponytail etc. (to support conventional retrieval needs). The complete relevance judgments for each query topic were also generated and will be released with the topics in the dataset addendum ².

The main reason for developing broad and narrow focus user queries (specifically for the dynamic approach) is to provide the access to almost all of the user’s daily living activities (regular to occasional) for the retrieval task. Moreover in the static approach to event segmentation, the dataset is already segmented into human oriented regular (known) activities that is fixed in nature so these query topics do not matter for the static approach; they are only required for the dynamic approach evaluation.

C. Evaluation Methodology

Identifying a robust evaluation mechanism for the minor variations in the exact boundary definitions (for example, few minutes either side) is an important consideration for lifelog segmentation evaluation. Sliding window techniques are still the predominant tool for activity boundary detection and segmentation, as discussed in Doherty et al. [11], [17] and as used in our baseline approach. Thus, we implemented

²LSC2018 Dataset available at: <http://lsc.dcu.ie>. Additional annotations to be released in July 2019.

the sliding window algorithm (explained in [17]) to evaluate the proposed static and dynamic segmentation approaches. Consequently, a number of different window frames f (where $f = 1$ to 10) have been employed to observe the accuracy of detected activity boundaries. We split and slide over the whole sequence of images into a set of sliding windows ($\pm [1$ to 10]) that provide a necessary degree of flexibility in measurement process. It was our belief that bigger the frame size has to provide higher score. Through experimentation, we observe that the accuracy measure get constant at particular frame size (where precision score reaches at 1 and trade-off between precision-recall value become constant) for the further frame sizes. Consequently, we found that the frame acquiring five minutes is the most appropriate size of the sliding window for evaluating both static and dynamic activity segmentation algorithms because the trade-off between precision and recall (at $f = 5$) is more consistent in terms of getting almost similar scores for both precision and recall (see Fig. 3), which is in line with the assumptions of Doherty et al. [17].

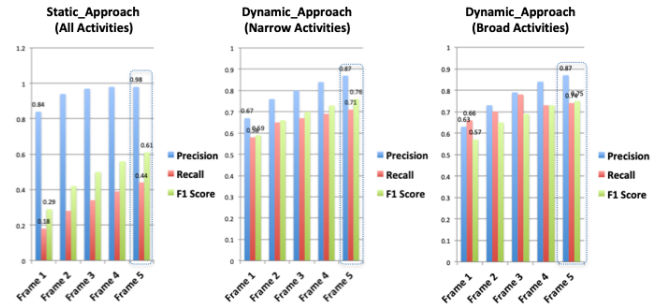


Fig. 3. Identifying an optimal window size in sliding window algorithm.

If the system defined boundary is found within 5 minutes of the human judgment, then it is considered to be accurate and we fixed that boundary as a segmentation point. This methodology used the standard evaluation measurements such as precision, recall, and f1 score. Here, we calculated the precision in terms of true positives where negative effect of over segmentation is ignored, recall represents the fraction of relevant documents that are successfully retrieved by the algorithm and f1 score represents harmonic mean of precision and recall.

D. Results

We compare both static and dynamic activity segmentation approaches for all 12 broad and 12 narrow focus queries (discussed in section IV(B)). We show that ranked list of topic specific activity boundaries (in the dynamic approach) would provide a flexible mechanism to access user information, rather extracting a ranked list of all activities performed in a day (in static approach).

1) *Baseline Static Segmentation Approach*: Detecting the sequential activity boundaries performed in a particular day is prime objective of our static segmentation approach. We revised and implemented the existing evaluation methodology based on [11], [17] and get the improved score of precision

TABLE I
ACCURACY SCORES FOR EACH ACTIVITY IN DYNAMIC SEGMENTATION APPROACH.

| Broad Topic User Queries | | | | Narrow Topic User Queries | | | |
|--------------------------|-------------|-------------|-------------|---------------------------|-------------|-------------|-------------|
| Topics | Precision | Recall | F1 score | Topics | Precision | Recall | F1 score |
| working on laptop | 0.94 | 0.82 | 0.85 | walking to airplane | 0.80 | 0.63 | 0.70 |
| work meeting | 0.74 | 0.70 | 0.70 | walking at work | 0.97 | 0.77 | 0.84 |
| walking | 0.91 | 0.96 | 0.92 | brainstorming | 0.56 | 0.72 | 0.63 |
| relaxing at home | 0.74 | 0.67 | 0.67 | browsing items | 0.96 | 0.85 | 0.88 |
| socializing | 0.79 | 0.79 | 0.76 | packing | 1 | 0.40 | 0.55 |
| travelling in airplane | 0.58 | 0.95 | 0.70 | waiting for train | 1 | 0.53 | 0.68 |
| cooking | 0.99 | 0.53 | 0.65 | gardening | 0.75 | 0.64 | 0.69 |
| cleaning | 1 | 0.44 | 0.57 | writing on paper | 1 | 0.95 | 0.97 |
| driving | 0.99 | 0.51 | 0.64 | saturday morning coffee | 0.58 | 0.73 | 0.64 |
| retail shopping | 0.95 | 0.77 | 0.84 | fruit bowl | 1 | 0.81 | 0.89 |
| reading paper | 0.94 | 0.89 | 0.91 | eating apple | 1 | 0.77 | 0.87 |
| watching TV | 0.82 | 0.83 | 0.79 | ponytail | 0.57 | 0.73 | 0.63 |
| Average | 0.87 | 0.74 | 0.75 | Average | 0.85 | 0.71 | 0.75 |

(0.98), recall (0.44) and f1 score (0.61) as compared to a baseline approach (as illustrated in Table II). We observe that the recall value is less than half of the precision value. This is due to the nature of static algorithm where all activity boundaries are predefined and fixed. Usually, the activity boundaries defined by the users are fewer in number (around 25-30 per day) as compared to activity boundaries defined by our proposed segmentation algorithm (around 80-90 per day). This obviously results in the increment of precision value (where almost all activity boundaries get detected accurately while evaluating segmentation algorithm) but reduces the recall value because of missing a large number of inaccurately detected activity boundaries.

2) *Dynamic Segmentation Approach*: By implementing dynamic segmentation algorithm, we get the highest score of precision (0.85), recall (0.71), and f1 score (0.75) from narrow focus user queries and get highest score of precision (0.87), recall (0.74) and f1 score(0.75) for broad focus user queries (see Table II). Therefore, we can observe the improvement in both precision and recall scores using our proposed evaluation algorithm as compared to our previous work [16], [20]. Observing the accuracy for each specific activity (illustrated in Table I), we get the highest score (in terms of f1 score) for the broad type activity such as walking on the lovely day (0.92) with search criteria *Keywords = walking, outdoor, sunny, noon*. Similarly, we get the highest score (in terms of f1 score) for the narrow type activity such as writing on paper (0.97) with search criteria *Keywords = paper, work, pen*. We observe that some activities such as cleaning (0.57), packing a suitcase(0.55), brainstorming (0.63) and ponytail (0.63) are still performing poorly (see Table I). It is our belief that, the limited frequency of occurrence of such activities in the corpus could be the reason for this.

3) *Overall Best Segmentation Approach*: As we can observe from Table II that the overall accuracy (in terms of average f1 score) for dynamic approach (0.75) outperforms the average f1 score for static approach (0.61). However, we still need to explore the specific query based measure

of improvement (for dynamic approach) against conventional static approach. Thus through experimentation, we compared both approaches and found the measure of improvement for the dynamic approach over static approach (see Table III).

TABLE II
PERFORMANCE OF STATIC AND DYNAMIC SEGMENTATION APPROACHES.

| Scores/Approach | Precision | Recall | F1 Score |
|-------------------------|-----------|--------|----------|
| <i>Static</i> | 0.98 | 0.44 | 0.61 |
| <i>Dynamic (Narrow)</i> | 0.85 | 0.71 | 0.75 |
| <i>Dynamic (Broad)</i> | 0.87 | 0.74 | 0.75 |

4) *Improvement over Static Segmentation Approach*: We deployed a new strategy that compares and generate a similarity vector based on the resulted activity boundaries from both approaches. Thus, we retraced the topic specific activity boundaries detected by dynamic approach to the activity boundaries detected by static approach. We get the reasonable level of improvement for (each activity in) dynamic segmentation approach over static segmentation approach i.e. 0.39 for broad focus and 0.26 for narrow focus user queries (see Table III).

TABLE III
DYNAMIC SEGMENTATION APPROACH OUTPERFORMS STATIC SEGMENTATION APPROACH.

| Improvement Over Static Segmentation Approach | | | |
|---|-------------|-------------------------|-------------|
| Broad Topics | Improvement | Narrow Topics | Improvement |
| working on laptop | 0.74 | walking to airplane | 0.68 |
| work meeting | 0.56 | walking at work | 0.60 |
| walking | 0.56 | brainstorming | 0.36 |
| relaxing at home | 0.56 | browsing items | 0.23 |
| socializing | 0.39 | packing | 0.22 |
| travelling in airplane | 0.37 | waiting for train | 0.22 |
| cooking | 0.36 | gardening | 0.19 |
| cleaning | 0.32 | writing on paper | 0.14 |
| driving | 0.31 | saturday morning coffee | 0.13 |
| retail shopping | 0.20 | looking at fruit bowl | 0.12 |
| reading paper | 0.16 | eating apple | 0.10 |
| watching TV | 0.12 | ponytail | 0.12 |
| Average | 0.39 | Average | 0.26 |

V. DISCUSSION

Through this experimental study, we found that the dynamic segmentation approach outperforms static segmentation approach. Although, we observe that mostly narrow type activities such as Saturday morning coffee (0.13), looking at fruit bowl (0.12), having conversation with the person who has ponytail (0.12), eating apple (0.10), and two broad type activities reading paper (0.16), and watching tv (0.12) proved to have less improvement over static approach (see Table III). We can identify a number of reasons for this, the narrow type user queries are more specific to access user information which might be of short (one or a few minutes) duration. Thus, it makes the algorithm produce lower results (in terms of recall) to match the activity boundaries for such a small time stamp (activity specific) from huge corpus (all activities/day). It is our inference that the query topics should not be too specific so that it could only retrieve one or few minutes of relevance judgments. We intend to explore this idea more further that could provide useful insights to research community.

VI. CONCLUSION

In this study, we implemented our a static segmentation approach and dynamic segmentation approach to identify the better approach for activity boundary detection and segmentation. The segmentation model is evaluated using the reusable LSC2018 lifelog test collection with manual annotations (included in feature vector) and additional user query topics. We revised the existing sliding window approach [17] by employing different sizes frames. We suggest that, there is a significant scope for building better and enhanced version of activity based ad-hoc search engines. In future, we plan to extend this work by employing better moment metadata and further revise the retrieval approach so as to reduce the size of the sliding window to less than five.

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] J. Gemmell, G. Bell and R. Lueder (2006), MyLifeBits: a personal database for everything. In: *Communications of the ACM*: 88-95.
- [2] A. R. Doherty, A. F. Smeaton, K. Lee and D. P. W. Ellis (2007). Multimodal segmentation of lifelog data. In: *Large Scale Semantic Access to Content (Text, Image, Video, and Sound)*: 21-38.
- [3] A. R. Doherty, C. J. A. Moulin and A. F. Smeaton (2011) Automatically assisting human memory: A SenseCam browser, *Memory*, 19:7, 785-795.
- [4] S. Yamamoto, T. Nishimura, Y. Takimoto, T. Inoue and H Toda (2017). PBG at the NTCIR-13 Lifelog-2 LAT, LSAT, and LEST Tasks.
- [5] C. Gurrin, H. Joho, F. Hopfgartner, L. Zhou, R. Gupta, R. Albatat and D. T.Dang Nguyen (2017). Overview of ntcir-13 lifelog-2 task. The Thirteenth NTCIR conference (NTCIR-13).
- [6] A. G. d. Molino, J.H. Lim and A.H. Tan (2018). Predicting Visual Context for Unsupervised Event Segmentation in Continuous Photo-streams. In: *ACM Multimedia Conference (MM '18)*.
- [7] H. Miller, P. Clough, T. Deselaers and B. Caputo (2010), *ImageCLEF: Experimental Evaluation in Visual Information Retrieval*, Springer Publishing Company, Incorporated.

- [8] 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/>.
- [9] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadar-rana and T. Darrell (2014). Caffe: Convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093.14.
- [10] H. Fang, S. Gupta, F. Iandola, R. Srivastava, L. Deng, P. Dolla, J. Gao, X. He, M. Mitchell, J. Platt, L. Zitnick and G. Zweig (2015). From captions to visual concepts and back. *IEEE Institute of Electrical and Electronics Engineers*.
- [11] R. Gupta and C. Gurrin (2018) Approaches for Event Segmentation of Visual Lifelog Data. In: Schoeffmann K. et al. (eds) *MultiMedia Modeling, MMM 2018. Lecture Notes in Computer Science*, vol 10704. Springer, Cham.
- [12] E. Talavera, M. Dimiccoli, M. Bolaos, M. Aghaei and P. Radeva (2015). R-clustering for egocentric video segmentation. In *Iberian Conference on Pattern Recognition and Image Analysis: 327-336*. Springer International Publishing.
- [13] M. Dimiccoli, M. Bolaos, E. Talavera, M. Aghaei, S. G. Nikolov and P. Radeva (2015). SR-Clustering: Semantic Regularized Clustering for Egocentric Photo Streams Segmentation. In *Computer Vision and Image Understanding Journal (CVIU)*.
- [14] C. Gurrin, K. Schoeffmann, H. Joho, A. Leibetseder, L. Zhou, A. Duane, D.T. Dang-Nguyen, M. Riegler, L. Piras, M-T. Tran, J. Lokoč and W. Hürst (2019). Comparing Approaches to Interactive Lifelog Search at the Lifelog Search Challenge (LSC2018). *ITE Transactions on Media Technology and Applications*: 46-59.
- [15] S. Abigail and W. Steve (2010). Beyond total capture: A constructive critique of lifelogging: 70-77.
- [16] R. Gupta and C. Gurrin (2019). "Considering manual annotations in dynamic segmentation of multimodal lifelog data." *ARDUOUS'19-3rd International Workshop on Annotation of useR Data for Ubiquitous Systems*. IEEE European Union.
- [17] A. R. Doherty and A. F. Smeaton. Automatically Segmenting LifeLog Data into Events. 2008 Ninth International Workshop on Image Analysis for Multimedia Interactive Services (2008): 20-23.
- [18] D. Byrne, B. Lavelle, A. R. Doherty, G. J. Jones and A. F. Smeaton (2007). Using bluetooth and GPS metadata to measure event similarity in SenseCam Images.
- [19] M. Bolaos, M. Garolera and P. Radeva (2014) Video Segmentation of Life-Logging Videos (2014). In: Perales F.J., Santos-Victor J. (eds) *Articulated Motion and Deformable Objects. Lecture Notes in Computer Science*, vol 8563. Springer, Cham.
- [20] R. Gupta and C. Gurrin (2019) Dynamic Segmentation and Retrieval of Daily Life Activities from Multimodal Lifelogs. In Press: Perner P. (eds) *Machine Learning and Data Mining in Pattern Recognition. MLDM 2019. Lecture Notes in Computer Science*. Springer International Publishing, 'unpublished'.
- [21] G. Bell and J. Gemmell (2007). A digital life. *Scientific American*.
- [22] W. Lin and A.G Hauptmann (2006). Structuring continuous video recordings of everyday life using time-constrained clustering.
- [23] 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 (2001). Human brain activity time-locked to perceptual event boundaries. In: *Nature neuroscience*: 651-655.
- [24] R. Gupta, Considering documents in lifelog information retrieval (2018), *ICMR18. In: International Conference on Multimedia Retrieval*: 497-500.