Overview of Lifelogging: Current Challenges and Advances

AMEL KSIBI[1], ALA SALEH D. ALLUHAIDAN[1], (Member, IEEE), AMINA SALHI[1], AND SAHAR A. EL-RAHMAN[2,3], (Senior Member, IEEE)

[1]Information Systems Department, College of Computer and Information Science, Princess Nourah Bint Abdulrahman University, Riyadh 11564, Saudi Arabia
[2]Computer Sciences Department, College of Computer and Information Science, Princess Nourah Bint Abdulrahman University, Riyadh 11564, Saudi Arabia
[3]Electrical Engineering Department, Faculty of Engineering-Shoubra, Benha University, Cairo 13511, Egypt

Corresponding author: Ala Saleh D. Alluhaidan (asalluhaidan@pnu.edu.sa)

ABSTRACT Lifelogging is the process of digital tracking of person's daily experiences for a variety of purposes. In recent years, lifelogging has become an increasingly popular area of research due to not only the growing demands from many applications, such as wellbeing, entertainment, healthcare systems, and intelligent environments, but also to the advances in device technologies that offer the promise to record and store large volumes of personal data using inexpensive tools. However, getting insights from egocentric experience using huge deluge of unlabeled and unstructured data continues to pose major challenges. A large number of research have been conducted in recent years to cover these challenges but there is still a lack of studies that provide a comprehensive survey of the available literature, while most of the existing lifelogging surveys generally focus on only one aspect. This review highlights the advances of state-of-the-art in lifelogging from different angles, including its research history, current applications, activity recognition techniques, moment retrieval, storytelling, privacy and security issues, as well as challenges and future research trends.

INDEX TERMS Activity recognition, challenges, lifelogging, moment retrieval, security, storytelling, trends, privacy.

I. INTRODUCTION

The topic of lifelogging has been around for a long time as it has started as records of events in diary published on social media like Facebook, Twitter, and Instagram. With the rapid increase in wearable technologies, over recent years, and advancement in storage, cloud services, sensing technology, and location awareness, recording events became easier, faster and more efficient.

Lifelogging, generally, is defined as recording personal life including daily experiences via wearable sensors such as accelerometers, belts, FitBit, cameras, and others [1]. For example, there is an increased number of people who track their physical activity using wearable sensors to analyze their performance in variety of tasks. Lifelogging became an essential service for our wellbeing. This was a result of

The associate editor coordinating the review of this manuscript and approving it for publication was Senthil Kumar .

convergence of technologies to enhance our lifestyle. For example, people are used to putting their wearable cameras in personal spaces (such as home, cars, garden,...) to record unforseen events such as accidents or assaults, or to monitor safety-critical situations for healthcare assistance. Recent developments in pervasive computing research have repositioned SenseCam as an aid for human memory and monitoring.

The small devices such as wearable cameras automatically and passively record daily activities. Data recorded at specific moments using video cameras provide deeper overview of daily activities. Moreover, data acquired over long periods of time are potential for getting insights from behavior patterns. Indeed, Lifelogging analysis could be useful to protect against diseases linked to unhealthy lifestyles such as obesity, depression,…etc. Additionally, this type of analysis could also helpful to inhibit cognitive and functional drop in elderly people.

INTRODUCTION

History & Definition

Aspects of Lifelogging

Objective of Research

Paper Organization

APPLICATIONS, CHALLENGES, AND TRENDS

Applications:
- Healthcare & Wellness
- Quality of Life

Challenges:
- Manual Entry
- Synthesize Data

Trends:
- Tracking and Monitoring
- Behavior Tracking

ACTIVITY RECOGNITION

Vision Based   Hybrid   Sensor Based

LIFELOG MOMENT RETRIEVAL

Query Topics Understanding   Automatic vs Interactive Retrieval Process

Search Results Diversification

STORYTELLING

Egocentric Summarization

Object Driven Approach   Deep Learning

PRIVACY AND SECURITY CHALLENGES

Technology   Logs, personal Data.   Geolocation
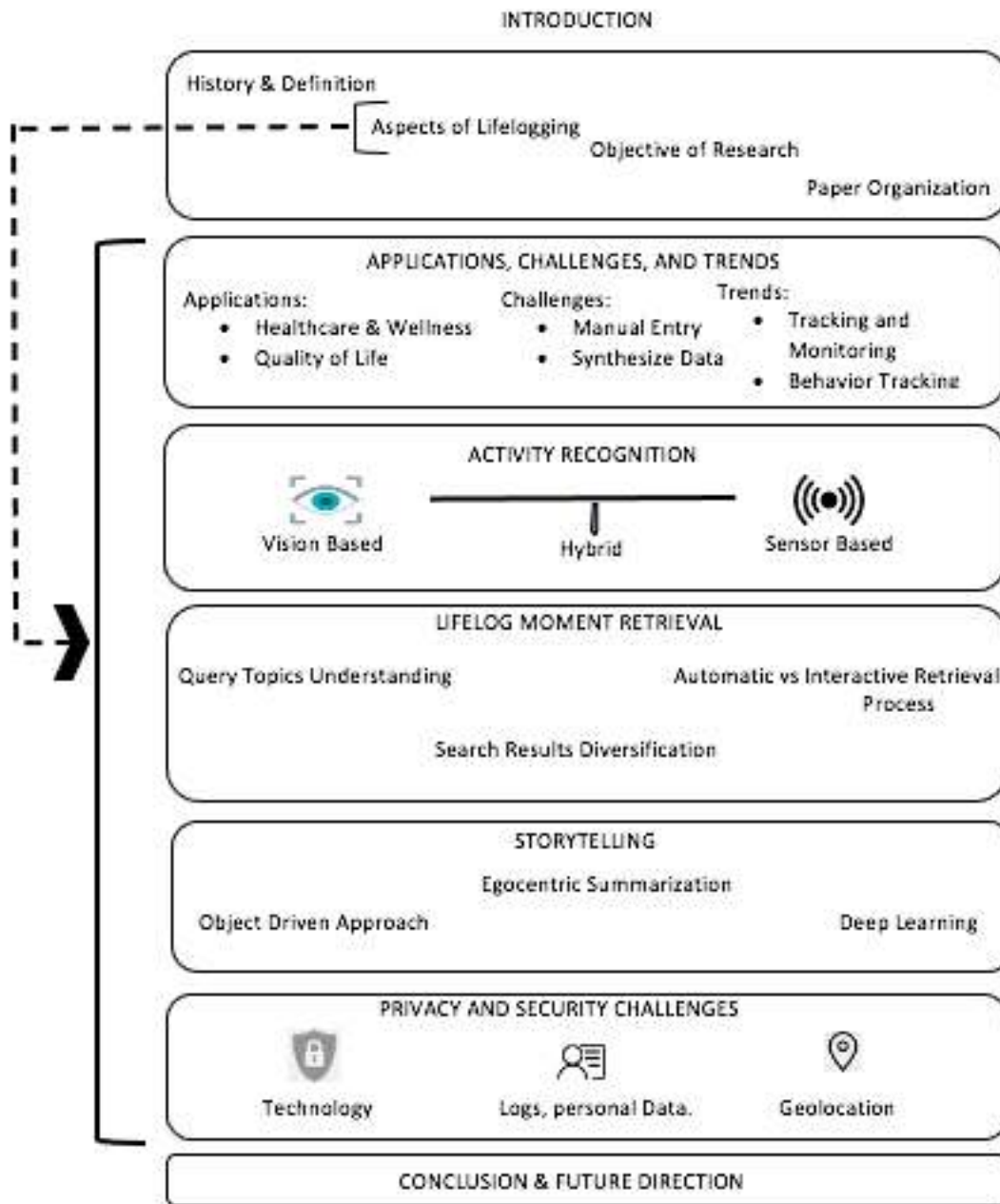

CONCLUSION & FUTURE DIRECTION

FIGURE 1. Aspects of lifelogging.

Lifelogging stories are captured automatically using egocentric photo streams or video cameras. Some hybrid approaches use both photo-graphic and video cameras [2]. which yield to a massive amount of unlabeled collections that require specific tools to understand the semantics of these images and videos. Also, dealing with free motion of cameras and the changes of lighting conditions along with image content make analysis techniques more challenging. As such, object and activity recognition algorithms should also be able to process huge number of images and videos with variety of objects. Additionally, computational resources needed to handle this issue should be reliable and consistent [2].

Another important dimension of lifelogging is ethics and security while managing data, specifically, for health-related application using wearable cameras [3], [4].

This review aims to provide an inclusive coverage of lifelogging, starting with presenting the diverse domains of lifelogging applications, then, detailing the importance of activity recognition, following the challenges of moment retrieval and storytelling and concluding with highlights of lifelogging security and privacy issues. Thus far, most of the lifelogging research has focused mainly on one aspect [5], [6] while we provide different views to the topic which is the contribution of this research (Figure1).

Thus, this paper will contribute as a comprehensive reference to help researchers understanding the lifelogging aspects and to provide them possible future directions.

II. LIFELOGGING APPLICATIONS, CHALLENGES AND TRENDS

Lifelogggging has been utilized and incorporated in many fields. Heathcare, wellbeing, and quality of life. For instance, lifelogging has been used to monitor behavior change for weight management in obese patients. The study [1] presents a mobile app for obese adults to monitor behavior focusing on (technical effectiveness, user efficiency, and user satisfaction). "Participants were asked to complete eight tasks for evaluating the technical effectiveness of the app." [1] Timing was used to evaluate user efficiency. For user satisfaction, participants were asked to complete the System Usability Scale (SUS); sample size of 50 adults (14 men and 36 women, aged 20-59 years) was targeted. The app collects behavioral information through a questionnaire that includes questions about diet, sleep time, and stress. Results show that the app has a satisfactory technical effectiveness, user efficiency, and user satisfaction with further needed clinical efficacy evaluation.

A theoretical framework presented by hypothesis to explain the value of SenseCam for memory retrieval in [2]. A SenseCam, a wearable camera, is used for image capturing to reinstates thoughts, feelings, and sensory information. In memory impairment, Dubourg *et al.*, [2] propose that the environmental support can be employed via SenseCam for memory retrieval to further improve episodic information retrieval.

Japan is imposing a new approach for monitoring patients using lifelogs in order to improve the health care, especially for elderly. A lifelog records person's activity and can be used to predict a lifestyle-related disease. This prediction could be helpful for the healthcare of the elderly. Additionally, building a self-recording platform integrated with the medical platform is a convenient way to have all data in one place. Such a system can be used to send personalized health advices [7].

A framework specialized for dementia care based on lifelogging monitoring with activity recognition from egocentric vision and semantic context-enrichment is presented in [8]. Within this framework, multimodal egocentric data are collected from wearable bracelet and the accelerometer to give more accurate description of patient's health state. Specifically, mechanical variables that include fine motion, such as jerk, enhances the recognition accuracy of activities. Furthermore, for building interoperable activity graphs using Semantic Web technologies, Meditskos *et al.*, [8] present a framework for semantic activity representation and interpretation. Results show that the proposed system was successful and efficiently personalized with specific activity models. By using lifelogging applications, not only healthcare givers were able to support interventions, but also end-users felt more safe and confident [8].

In an attempt to evaluate a wearable lifelogging camera in a sample of older adults diagnosed with mild cognitive impairment (MCI), collective data such as a self-report questionnaire, images taken by users, and chains of focus group discussions were gathered. Results show good acceptance and usage of the camera along with adequate number of images taken daily. Factors measured were perceived severity and ease of use. Privacy concerns are overlooked, focusing on potential benefits for memory [9].

Lifelogging has been used to raise awareness regarding quality of life with wearable trackers, smartphone sensors, and manual entry to collect indicators. A general infrastructure for collecting and processing life-logs, and how the quality of life indicators are calculated with GUI of life Meter are described in [6]. Findings indicate good usability of such application pointing to how it supports users in raising their awareness of monitoring quality of life [5].

Similar to SenseCam, an attempt for lifelogging was used for monitoring of dietary intake. Of course, entry here depends on manual logging. As an example, the DietSense project uses a mobile phone with a camera to automatically collect pictures of the wearer's day. The images are used as log of the wearer's mealtimes and are further used to analyze the diet intake for obtaining feedback and improving diet choices [6]. Also, Meditskos *et al.*, [8] used foodlogs in order to assess food balance for personal dietary monitoring and then Aizawa [10] suggests making societal donations through multimedia food log.

Some of the lifelogging applications require data from multiple sources to give more insight and meaningful result. Therefore, fusing multiple sensors data is needed and it imposes a challenging task as it requires data cleaning, alignment, and temporal normalization [6]. Additionally, manual entry of some data in different applications can increase the complexity of data processing as well as application's interface of the app. More challenges are discussed later on in sections IV. Lifelog Moment Retrieval and V. Storytelling. Lifelogging applications were extended to smoking cessation [11], since it is difficult to automatically sense smoking without wearable sensor or manual entry. Similarly, sleep monitoring is used for collecting time and quality of sleep. In previous examples, lifelogging were used for health promotion and to gain extra awareness of life activities. Other applications for lifelogging include: self-reporting errors in travel behavior, physical activities, sedentary behaviors, and forgotten calories [6]. Table 1 summarizes the different domains where lifelogging applications are implemented.

With data collected using lifelogging applications, methods for extracting, processing, and summarizing need to be innovative. Next section is a review of current algorithms used to tackle each dimension of lifelogging.

III. ACTIVITY RECOGNITION

In recent years, sensors and hardware technology has reached the point where it became possible to record and store unlimited range of information that is generated in a lifetime by

TABLE 1. Summary of conditions and targeted domains by lifelogging applications.

No	Domain	Year	Article	Condition	Ref
1	Diet	2019	Development and Usability of a Life-Logging Behavior Monitoring Application for Obese Patients	Obesity	[1]
2	Healthcare	2019	Super-aged society: Constructing an integrated information platform of self-recording lifelogs and medical records to support health care in Japan	General healthcare for elders	[6]
3	Healthcare	2019	Acceptability of a lifelogging wearable camera in older adults with mild cognitive impairment	Older adults with mild cognitive impairment	[10]
4	Healthcare	2018	Multi-modal activity recognition from egocentric vision, semantic enrichment and lifelogging applications for the care of dementia	Dementia	[12]
5	Healthcare	2016	Sense- Cam: A new tool for memory rehabilitation?	Memory rehabilitation	[2]
6	Healthcare	2011	The Potential of Sensor-Based Monitoring as a Tool for Health Care	Smoking cessation	[11]
7	Quality of life	2017	Increasing Quality of Life Aware-ness with Life-Logging	Quality of life awareness	[5]
8	Healthcare and wellness	2014	LifeLogging: Personal Big Data, Foundations and Trends in Information Retrieval	Sleep monitoring, self- reporting errors in travel behavior, physical activities, sedentary behaviors, and forgotten calories	[6]
9	Healthcare and wellness	2013	Diverse Applications from Self-Monitoring to Social Contributions	Diet intake, donation	[10]

a single person [13]. However, the possibilities of analyzing them in an automated way are rather limited, and until now, computer vision systems are far from being comparable to human vision. Automatic human behavior analyzing and understanding remains a complex subject, and it has been for a long time one of the main goals of artificial intelligence practitioners [14].

Human Activity recognition can be defined as the ability to detect human gesture or motion based on data received from different sensors, and then interpret it into well-defined activity or action. Sensors can be cameras, wearable sensors, external sensors deployed in the environment, among others [15]. Acquired data provide considerable potential for knowledge mining concerning people performance during their lives, hence, they open up new opportunities for many potential applications in various fields including healthcare, remote monitoring, ambient intelligence, smart homes, security and surveillance, and human-computer interaction [16], [17].

The existing literature in human activity recognition could be divided into three main categories: vision-based, sensor-based, and hybrid system-based approaches [18]. Table 2 shows which type of data, imitation, and algorithm of each approach.

A. VISION BASED APPROACHES
Vision-based approaches consist of using camera to capture data, hence providing rich contextual and environmental information about the performed behavior [19].Due to its ease of use, computer vision techniques has been widely applied, to recognize the different activities from captured data with satisfactory results. However, it still a challenging task due to numerous problems, such as privacy, light dependency, occlusion, background clutter, and camera motion. In [20]

authors presented a detailed survey on existing methods and their abilities for handling with the above-mentioned challenges.

A typical activity recognition task from video stream or still images generally involves two major steps [21], action representation and action classification. The goal of action representation is to convert an action into a series of feature vectors; these features should be representative distinctive and invariant to improve the recognition performance [22]. Then, in action classification, the activity category will be inferred based upon the input feature vector [23]. With the emergence of deep learning, the above-mentioned two steps are merged into a unified end-to-end trainable framework, where action features can automatically learned from acquired data [24].

Action representation methods are summarized into global representation, local representation, and recent depth-based representation. The earliest studies attempted to extract global descriptors from acquired videos or images and encode them as a whole features. Bobick and Davis [25] presented Motion Energy Image (MEI) and Motion History Image (MHI) framework to encode dynamic human motion into a single image. However, these methods are sensitive to viewpoint changes. Weinland *et al.* [26] propounded the 3D motion history volume (MHV) to overcome the viewpoint dependency. In [27] Kumari and Mitra proposed a discrete Fourier transforms (DFT) based approach to obtain information about the geometric structure in the spatial domain. Furthermore, in [28] Tasweer *et al.* used a blocked discrete cosine transform (DCT) from motion history image (HMI) to extract global features.

Different from global representations, Local representations only focus on specific local regions having salient

TABLE 2. Activity recognition comparison.

No	Year	Modality	Features representation (global/local/depth)	Classification algorithm	Dataset	Recognition accuracy
[22]	2017	RGB	Depth, using a combination of spatiotemporal interest points, and dense trajectory features with bag-of-words model	Deep Sequential Context Network	(a) UCF101 dataset; (b) Sports-1M dataset; (c) BIT dataset	(a) 81.31%; (b) 68.62%; (c) 72.97%
[24]	2017	RGB	Depth, using end-to-end learning spatiotemporal features	deep residual networks (ResNets) model	(a) UCF101 dataset; (b) HMDB51 dataset	(a) 94.2%; (b) 68.9%
[26]	2006	RGB	Global, using 3D motion history volumes	Principal component analysis (PCA) with Mahalanobis distance	IXMAS dataset	93.33%
[27]	2011	RGB	Global, using discrete fourier transform (DFT) on silhouettes	K-nearest neighbor	(a) MuHaVi dataset; (b) DA-IICT dataset	(a) 96%; (b) 82.66%
[28]	2015	RGB	Global, using blocked discrete cosine transform (DCT) features from motion history image (HMI)	K-nearest neighbor	KTH dataset Weizmann dataset	89.9% precision, 88.7% recall
[31]	2013	RGB	Local, using improved dense trajectory with HOG, HOF, MBH	SVM	(a) Hollywood2 dataset; (b) HMDB51 dataset; (c) Olympic Sports dataset; (d) UCF50 dataset	(a) 64.3%; (b) 57.2%; (c) 91.1%; (d) 91.2%
[32]	2014	Depth motion map (DMM)	Depth, using low-level poly-normal assembled from local neighboring hyper-surface normal and are then aggregated by Super Normal Vector	Linear classifier	(a) Microsoft Research (MSR) Action 3D dataset; (b) MSR Gesture 3D dataset; (c) MSR Action Pairs 3D dataset; (d) MSR DailyActivity 3D dataset	(a) 93.45%; (b) 94.74%; (c) 100%; (d) 86.25%
[34]	2017	Depth motion map (DMM)	Depth, using multi-features extracted from human body silhouettes with joints information	Hidden Markov model	(a) Online self-annotated dataset; (b) MSR Daily Activity3D dataset; (c) MSR Action 3D dataset	(a) 71.6%; (b) 92.2%; (c) 93.1%
[35]	2018	Depth motion image (DMI)	Depth	convolutional neural networks	(a) MSRAction3D; (b) UTD-MAHD; (c) MAD dataset	91.86%
[37]	2008	RGB	Local, using space-time interest points (STIP) with HOG, HOF are encoded with BoVW	SVM	(a) KTH dataset; (b) Hollywood dataset	(a) 91.8%; (b) 38.39%
[38]	2008	RGB	Global, 23 degrees body model	Hierarchical variable transition HMM (HVT-HMM)	(a) Weizmann dataset; (b) gesture dataset	(a) 100%; (b) 90.6%
[39]	2010	RGB	Local, YIQ color model for skin pixels; histogram-based color model for face region; optical flow for tracking of hand motion	Dynamic Bayesian network	Collected dataset: 498 instances: (a) 10 gestures × 7 subjects × 7 times (isolated gesture); (b) 8 longer videos contain 50 gestures (continuous gestures)	(a) 99.59% accuracy; (b) 84% recall and 80.77% precision
[45]	2018	Frequency-domain	spectrogram-based feature extraction using short Fourier transform (STFT)	deep long term short term memory (LSTM) neural network architecture	(a) Collected dataset; (b) UCI HAR dataset	(a) 88.7% (b) 88.87%

motion information which are determined by interest point detectors, and thus inherently overcome the problem in global approaches. Furthermore, extracted features are more stable, and more robust to the corresponding transformations, occlusion and appearance variation [29]. The majority of local features extraction and representation methods was based on space-time interest points [30], and motion trajectory [31].

Recently, the emergence of depth cameras open the possibility to take advantage of depth maps that contain additional depth coordinates comparing to conventional RGB images and are able to capture color images sequences together with depth maps in real time [32]. Moreover, depth images are more robust to factors such as illumination, cluttered backgrounds, and occlusions [33]. To this end, various depth representation have been explored. For example, Jalal et al. [34] attempted to fuse spatiotemporal features in

RGB data with depth data. Farooq et al. [35] proposed to construct depth motion maps (DMM) and to add the motion energy for each view, and then they calculate the body part of the action (BPoA) by bounding box with an optimal window size for each DMM to get the action recognition. Kamel et al. [36] proposed to use depth maps and posture data with convolutional neural networks for human action recognition. In the next stage, the classification algorithm will determine the action label or the action category. Different classifiers have been explored on the extracted features for activity recognition, such as support vector machine (SVM) [37], Markov model (HMM) [38] and dynamic Bayesian network [39], decision trees or random forest [40]. In recent years, the deep learning based human action recognition methods have become the major research direction and gradually replaced the traditional approaches [41].

B. SENSOR BASED APPROACHES

This approach is based on the use of various types of emerging sensor, such as accelerometer, to determine human behavior. Sensors can be wearable, attached directly or indirectly the actor body, or can be dense sensors, embedded to objects that constitute the environment [42]. The generated data can be regarded as a continuous time series of motion changes represented as parameter values, and various features will be then extracted from these data using statistical or structural approach. Finally, those features serve as inputs to a machine learning algorithm to recognize human's ongoing activity [43]. Some of the most common machine learning algorithms, which are used in human activity recognition: support vector machine (SVM) [44], Long Short-Term Memory Network (LSTM) [45], Random Forest [46], and Convolutional Neural Networks (CNNs) [47].

Recently, the advancement of deep learning makes it possible to perform automatic extraction of high-level and more meaningful feature, which is more suitable for complex activity recognition. Furthermore, when faced with a large amount of unlabeled data, the deep generative network structure is more suitable for unsupervised incremental learning. Since then, deep learning based methods have been widely adopted for the sensor-based activity recognition [48].

C. HYBRID SYSTEM-BASED APPROACHES

In real applications, no single sensor can handle all possible activities, and usually different sensors are combined to improve HAR systems and overcome the problems of using a single sensor [49]. Many works have shown an accuracy improvement using multiple sensors in recent years. For instance, Zdravevski *et al.* [50] proposed an enhanced and real-time multimodal sensor-based activity recognition system, it was based on the fusion of vision based sensors and inertial sensors using machine learning for health monitoring. Ozcan and Velipasalar [51] proposed to combine features extracted from photo stream acquired by a wearable camera with data acquired from an accelerometer to perform fall detection for elderly persons. [52], video and IMU data captured synchronously by Google Glass were used to recognize wearer activities. The fused data resulted in an average accuracy higher than the individual accuracies of video and sensor data respectively. Further details on human activity recognition using various sensor fusion are reported in [53], [54].

IV. LIFELOG MOMENT RETRIEVAL

Lifelogs represent rich repositories of individual's daily experiences. These sources of information requires proper tools for retrieving specific life moments. Hence, there is a compelling need for appropriate retrieval systems that accurately remind a lifelogger about past moments.

Rigorous comparative benchmarking tasks have been dealing with this issue such as Lifelog Semantic Access Task (LSAT) at NTCIR-14c [55], lifelog moment retrieval (LMRT) at the ImageCLEFlifelog 2019 [56] and the Lifelog Search Challenge (LSC) at ACM ICMR2019 [57], as shown in the table below. The task throughout these three competitions is similar. Given a topic of users' daily activity or event (e.g. Find the moment when a user was taking a train from the city to home) as a natural language query, the system should retrieve the most relevant and informative images of the moments from users lifelogs. Researchers consider this task as a tedious task due to different issues. The first issue is regarding the semantic gap. There is no direct connection between lifelog images and query topics representing events/activities. This issue has a direct impact on the relevance of the search results [58]. The second issue concerns the quality of images. Since photos are taken whilst on the move, there is a problem of blurriness. In fact, the blurred images will not provide enough information; yet reduce the efficiency of search performance due to the wasted computation time [59], [60]. The third issue is about images redundancy. Since the lifelogger may be in stationary situations during the day, duplicates photos tend to exist within the lifelogs [59]. The retrieval of such images is time consuming without any benefit.

According to the above-mentioned issues, relevance and diversity represent the major retrieval criteria to satisfy.

After an exhaustive search of the literature, three main areas can be distinguished in order to improve the relevance of retrieval: data augmentation based on pre-trained models, filtering blurred images and natural language processing for query topics understanding.

The advances of Deep learning in scalable Image Annotation lead to providing pre-trained models that extract effectively the visual concepts within different aspects such as attributes, objects, location, places. Therefore, most of research works tend to apply deep transfer learning [61] which leverages the outputs of different pre-trained models (VGG-19, Retinanet, InceptionV3, ResNet50, Faster RCNN [60], PlaceCNN [62], YOLOv3 [63]) over external resources such as ImageNet [64], Place365 [65], MS-COCO [66], Open Images [67], SUN [68] in order to enhance the initial concept annotation of lifelog images provided by the official competitions.

Instead of collecting the outputs of the pre-trained models, other research works opt for fine-tuning method by deriving dense feature vectors from the last layer of the pre-trained models to be re-used to retrain their models [69]–[73]. According to these works, fine-tuning method outperforms the use of pre-trained models.

For the noise of lifelog images, researchers focus on eliminating the blurred and uninformative images from the dataset. Min-Huan Fu *et al.* Proposed to apply lens calibration followed by blurriness and color homogeneity detection for pre-processing [58]. Soumyadeb *et al.* estimated Blur score using a Haar wavelet transform. Images below the threshold are pruned [59]. The UPB team [60] deals with the uninformativeness and blurriness. First, they ran a blur detection system that computes the variance of the Laplacian kernel for

TABLE 3. Summary of benchmarking tasks.

Task	Organizer	Task details	Dataset	Queries
Lifelog Semantic Access task (LSAT)	NTCIR-14	known-item search task applied over lifelog data: retrieval of a number of specific moments in a lifelogger's life in response to a query topic	multimodal lifelog data over 42 days by the two lifeloggers (images from wearable cameras, Biometrics Data, Human Activity Data, visual concepts)	24 topics
Lifelog moment retrieval (LMRT)	ImageCLEF	Retrieval of specific predefined activities in a lifelogger's life + the diversification of the selected moments with respect to the target scenario	General healthcare for rich multimodal dataset which consists of 29 days of data from one lifelogger (images from wearable cameras, visual concepts, semantic locations and activities based on sensor readings, biometrics information, music listening history, computer usage)	10 topics
Lifelog Search Challenge (LSC)	ACM ICMR2019	multimodal interactive lifelog search engine	one month of anonymized multimodal lifelog data (continuous capture images, locations, biometrics, information consumption/creation activities)	20 topics (10 for experts and another 10 for novice users)

each image in order to capture both motion blur and large homogeneous areas. Then, they applied restriction rules on meta-data in order to remove the uninformative images. The ZJUTCVR team [74] applied Laplacian filters to determine the blur as the variance of convolution result, and they calculated the proportion of subjects in each image in order to detect occluded images.

A major challenging task within lifelog retrieval is query topics understanding. A common approach is to apply natural language processing tools in order to tackle the complexity of queries. Most of reviewed works customized their solutions using this approach in different ways. Abdallah et al. [69] proposed an automatic retrieval system based on Long-Term-Short-Memory(LSTM) for query processing. They built labelled textual descriptions of queries moments. They converted these words into numeric vectors by training a word embedding. After that, they created and trained an LSTM network based on the sequences of word vectors in order to extract the relevant concepts representing each topic. The retrieval phase entails matching the extracted query concepts using LSTM with the file containing the image concepts. The UAPT team [63] proposed an automatic retrieval process that extracts relevant words from topics titles and narratives and matches them with the lifelog images using a word-embedding model trained on Google News dataset. The TUC MI team [75] processed automatically the query with Natural Language Processing techniques by introducing the concept of token vector, which has the same dimension as image/segment vector, and defining a formula to compare the similarity between image/segment and token vector.

In order to reduce the gap between the query topic and the visual concepts of images, Suzuki and Ikeda [76] applied the vector representation of words by training the word embedding with skip-gram model. Subsequently, the similarity is calculated between two bags-of-words for the query topic and visual concepts V representing the images. Since the image is worth more than one thousand words, they, also, proposed

to transform the topic query into a set of images and train a topic classifier using the convolutional neural network over a collection of web pictures representing query topics. This classifier determines a topic score, which is the relevance of an image B for the query topic. The global similarity between query topic Q and image B is calculated by the sum of cosine similarity in the embedded space and the topic score.

In order to cope with the weaknesses of the automatic retrieval process, user involvement process has been integrated aiming to enhance the quality of results through feedback mechanism. With the advent of the Lifelog Search Challenge (LSC), a number of interactive retrieval systems have been designed to support interactive retrieval from lifelogs [57]. LIFER2.0 is used as baseline system for Lifelog Moment Retrieval (LMRT) task in ImageCLEF2019. It is an interactive retrieval system based on faceted filtering and context browsing for gaining insights via simple interactions. As described in [72], the user query is submitted, as facets, in a criteria-matching engine based on similarity ranking. The LEMORE [73] system designs an interactive semantic engine that retrieves images based on their tags (high-level concepts) and temporal information. The engine combines matrix numeric processing and database queries for object and temporal tags, respectively. Chang et al. [58] proposed an interactive system that focuses on query expansion by suggesting top k similar concept words from the dataset. Due to the restriction of search in the embedding space, finding the top-similar concepts is computationally feasible by comparing the semantic similarity between concept words and query terms.

regarding the problem of results redundancy, Chang et al. [58] provide a filtering mechanism that removes similar images by calculating the nearest neighbors in the embedding space. In order to reduce the computational overhead, they built an offline KD-trees and applied clustering in the embedded space. Then, in the online phase, they applied the BM25 for document retrieval that measures the probabilistic relevance.

V. STORYTELLING

Storytelling is the art of telling stories. Currently, a new concept called Digital Storytelling is gaining popularity and aims to generate automatically digital stories using artificial intelligence. This concept is strongly related to visual Lifelogging, mainly with the spread of digital tools such as images, audio and video. In fact, Digital Storytelling targets insights into users through examining their Lifelogs and mining their daily activities and lifestyles. These insights enclose different areas such as healthcare, security, leisure, lifestyle and wellbeing. However, automatic building of digital stories from unstructured lifelogs poses a major challenge towards browsing and mining insights efficiently from a huge volume of unconstructed egocentric data.

Egocentric summarization techniques play an important role in overcoming this challenge by generating a concise and precise meaningful representation of lifelogs. By doing so, summarization can be seen as a support for the visualization, indexing, and browsing of historical events, with the least possible semantic loss (informativeness criteria) and the least information redundancy(representative criteria). Therefore, summarization is considered as multi-objective problem and can be defined as generating an optimized lifelog representation that maximizes the extracted information and minimizes the redundancy of information. According to these objectives, we found two categories of summarization approaches: informative selection approach and representative selection approach.

Recently, object-driven approaches become of great interest. Lee *et al.* [77] proposed a process that creates object-driven summaries for egocentric videos by selecting frames that reflect the key object-driven events. Indeed, they extracted region cues representing high-level of saliency in egocentric video, and then, they applied a regression method to predict the relative importance of any new region based on these cues. Guo *et al.* [78] proposed a method that focuses on extracting video shots which reflect high stable salience, discrimination and representativeness in order to generate compact storyboard summary [79]. Lu *et al.*, inspired from a work about studying links of news articles over time, defined a random walk-based metric that captures event connectivity beyond simple object co-occurrence, to provide a better sense of story [80]. Sun *et al.* identified the salient people and actions from videos 'in the wild' to depict a montage.

With the advances in Deep learning, there has been a great interest towards solving the problem of egocentric video summarization using unsupervised deep learning. In such orientation, the unsupervised video summarization is considered as a key frame selection problem. Shruti *et al.* extracted deep features using CNN and then they applied clustering algorithms to extract interesting keyframes [81]. Abhimanyu el al. addresses the problem of summarizing egocentric videos by applying deep features extraction and an optimal clustering approach (CSMIK K-means) through a combination of Integer Knapsack (IK) and CSM [82]. While Behrooz *et al.* applied long short-term memory network (LSTM) to learn a deep summarizer network via a generative adversarial framework for optimizing the frame selector [83]. In [84], Zhen appliyed key-frame selection based on multidimensional time series. They considered the sequential frame features as a composition of a set of one-dimensional time-series data, then they performed the CUSUM statistics to time series simultaneously for each dimension. following, several consecutive clips containing similar contents are obtained after segmentation using the calculated statistics.Finally, clustering process is performed for key frames selection in the obtained video clips.

Notably, Egocentric summarization has become a hot topic in new challenging international competitions that aim to develop benchmarks for summarizing egocentric lifelogging data and contribute to the improvement of summarization quality. In 2017, ImageCLEF2017 proposed Lifelog summarization (LST) task which aims to analyze all the lifelog images and summarize them according to specific requirements. The summary should be represented by 50 relevant and diverse images. ImageCLEF2018 organized Activities of Daily Living understanding (ADLT) task, which aimsgoal is to analyze the lifelog data for a given period of time and provide a summarization based on concepts describing the Activities of Daily Living and the contexts in which these activities take place. In 2019, NTCIR'14 published an exploratory task, Lifelog Insight Task (LIT), and the objective of this task is to gain insights into the lifelogger's daily life activities by providing an efficient/effective means of visualization of the data.

VI. PRIVACY AND SECURITY CHALLENGES

Lifelogging system is a framework for the daily recording of personal and sensitive data of individuals, and usually includes asset constrained smart objects (sensors) [3]. With that in mind, lifelogging is subject to some security threats, as each emergent technology is able to interfere with the life of each life-logger [3], [4]. There are many security risks of lifelogging in connecting smart world because the suitable security standards and protocols of smart objects are not yet mature. There are several security vulnerabilities related to smart objects, which resulted from the limited smart objects resources to support robust demonstrated security and cryptography techniques. Therefore, these techniques must be taken into consideration for overcoming the issues of privacy and trust, thereby preventing avoiding security risks especially with the fast development of lifelogging applications. Trustworthy and secure lifelogging involves security challenges, threats and risk of smart objects on the communication layer and on the users who login, sharing and exchange the private data using smart devices [4]. But sharing and exchange any type of personal information are threats to privacy, where infringements on privacy are considered as one of the key challenges of lifelogging [85].

In particular, lifelogging information needs trustworthy and powerful security considerations because it may contain very sensitive personal information such as communication
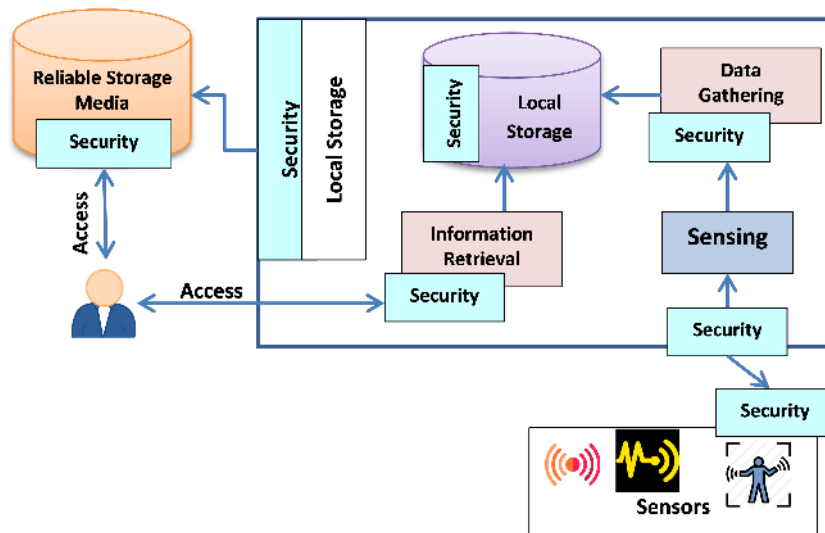
FIGURE 2. General lifelogging architecture with security requirements.

logs, location, biological information, etc. [4], [86]. Unfortunately lifelogging has a debatable history, like a DARPA's lifelogging project, that was cancelled due to the criticism of privacy indications of the system in 2004 [85], [86]. Also, regarding the lifelogging environment, the related risks of information security actually have substantial connotations on economy, society, privacy, and people's psychology [85], [87]. These security risks actually impact on individuals, industries, and governments [87]. The complexity of stockholders' behavior can also be evaluated in the social areas by utilizing social networks, messaging services, qualified self, and sharing of lifelogging information. A lot of those trends are in incomplete contradiction with the need of privacy based on well formulated human rights. Some individuals' information may be publicized where the governments consider it as confidential private data [88]. The number of security issues is encountered in the lifelogging trend that needs to be determined and solved to proceed into the future connected world.

To highlight the risks and the benefits of lifelogging [89], European Network and Information Security Agency (ENISA) in European Union (EU), [87] presented a scenario to study the challenges, risks, threats, and benefits of lifelogging in daily life. The main risks reported and detailed in [87] based on the individuals, industry, service providers, government, EU Institutions and regulators. ENISA addressed the risks related to lifelogging environment based on the stockholders; also ENISA reported a number of recommendations per stakeholder which they are addressed to. Petroulakis *et al.*, [89] detailed the lifelogging topics in smart environments, and described the security threats, interconnection issues, and suggested a lightweight framework to ensure privacy, security, trustworthy, and powerful lifelogging system based on the security attacks impact on energy consumption. The researchers applied several mitigation factors including AES-128 encryption, channel assignment, and power control for secure lifelogging system.

Fragkiadakis *et al.*, [3] developed a joint encryption approach and compression model based on the development in compressed sensing principles. Rawassizadeh and Tjoa [86] discussed the security related risks for lifelogging system and the risks of sharing lifelogging information in the social communities. The researchers presented a sharing model that can reduce the sharing ability of a lifelogging information object using the expiration time. Allen [90] specified two potential hazards of lifelogging that are pernicious surveillance and pernicious memory.

Generally, a lifelogging process consists of three phases and each phase requires particular security considerations. 1st phase is sensing the data from the customer environment with sensors, 2nd phase is gathering the sensed data, and 3rd phase allows the information browsing and retrieval from the user's lifelogging dataset. The general lifelogging architecture with Security Requirements as indicated in Figure2.

Lifelogging architecture parts that require to be secured are highlighted. Users need to define the collected information object and able to set configuration parameters of the sensors for example sensing interval. In the first phase, the lifelogging architecture is connected to sensors and reads the data of the sensor. Where some sensors need to be secure as they might need authentication and the confidential information in the sensor data needs to be secure using data encryption techniques through the data transferring from the sensor to the lifelogging architecture. So, the modules of dynamic security are provided for the sensing phase due to adding or removing the sensors dynamically to lifelogging architecture, which achieves a more flexible and scalable lifelog architecture. In the second phase, the sensed data in the lifelog object is collected to generate a lifelogging dataset of lifelogging records. Security modules should be considered by the developers through the gathering stage of data. In the third phase, the lifelog data is stored in storage devices which must be secured [86]. The most common techniques for securing the

data are cryptography and data encryption using different techniques such as RSA (Rivest Shamir Adleman), AES (Advanced Encryption Standard), ECC (Elliptic Curve Cryptography), and Data Encryption Standard (DES) [4].

VII. CONCLUSION AND FUTURE DIRECTION

Lifelogging has the potential to enrich individuals' life with right analysis. This review summarizes current applications of lifelogging, its' trends and challenges, presents analysis, focused on the progress made so far in this context of computer algorithms, from a storytelling perspective. A separate section was dedicated to cover the current techniques in activity recognition. Also, we discuss available literature around the privacy and security regarding lifelogging applications. Referring to applications, there is more potential usage of lifelogging within healthcare and education to enhance current practices on both domains. There is also a need to develop more algorithms suited to data obtained through photo cameras, in particular, for social interaction detection and analysis, as well as for activity and context recognition. With visual storytelling, semantics has to be reserved to extract egocentric data and summarization using ontologies. A case study would be a great research breakthrough to address those gaps.

A promising area of research, has not been explored yet, is rendering a visual to text translation in real-time using multi-modal techniques in order to provide a human-like description of incidents happened in an event with a richer context.

The risks of sharing lifelogging data and security recommendations to reduce risks are introduced. Security and privacy challenges indicate the shortage of suitable security protocols and mechanisms in connecting the smart world. Therefore, the main recommendation for stakeholders' providers in lifelogging domains is to follow a mechanism that comply with high privacy and security standards for protecting the lifelogging users.

This paper also shed light on the problem of moment retrieval from visual lifelogs. Specifically, we examine main issues regarding the quality of retrieved results. Then, we review state of the art solutions for these challenges. Explicitly, we emphasize the importance of both dataset filtering and natural language processing with transfer learning in reducing the semantic gap and improving the relevance of results. The essential role of digital storytelling with different approaches to deal with egocentric video summarization problem is also presented. From there we cover the strengths of object driven clustering and deep feature selection. Yet, there is a need to improve aspects about natural language semantics. As next steps, in view of our analysis, we plan to improve the quality of the query interpretation using advanced NLP algorithms.

REFERENCES

[1] J. Y. Lee, J. Y. Kim, S. J. You, Y. S. Kim, H. Y. Koo, J. H. Kim, S. Kim, J. H. Park, J. S. Han, S. Kil, H. Kim, Y. S. Yang, and K. M. Lee, "Development and usability of a life-logging behavior monitoring application for obese patients," *J. Obesity Metabolic Syndrome*, vol. 28, no. 3, pp. 194–202, Sep. 2019, doi: 10.7570/jomes.2019.28.3.194.

[2] L. Dubourg, A. R. Silva, C. Fitamen, C. J. A. Moulin, and C. Souchay, "SenseCam: A new tool for memory rehabilitation?" *Revue Neurologique*, vol. 172, no. 12, pp. 735–747, Dec. 2016, doi: 10.1016/j.neurol.2016.03.009.

[3] A. Fragkiadakis, I. Askoxylakis, and E. Tragos, "Secure and energy-efficient life-logging in wireless pervasive environments," in *Human Aspects of Information Security, Privacy, and Trust*, L. Marinos and I. Askoxylakis, Eds. Berlin, Germany: Springer, 2013, pp. 306–315.

[4] N. E. Petroulakis, E. Z. Tragos, A. G. Fragkiadakis, and G. Spanoudakis, "A lightweight framework for secure life-logging in smart environments," *Inf. Secur. Tech. Rep.*, vol. 17, no. 3, pp. 58–70, Feb. 2013, doi: 10.1016/j.istr.2012.10.005.

[5] F. Zini, M. Reinstadler, and F. Ricci, "Increasing quality of life awareness with life-logging," in *eHealth 360°* (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering), K. Giokas, L. Bokor, and F. Hopfgartner, Eds. Cham, Switzerland: Springer, 2017, pp. 282–293, doi: 10.1007/978-3-319-49655-9_36.

[6] C. Gurrin, A. F. Smeaton, and A. R. Doherty, "LifeLogging: Personal big data," *Found. Trends Inf. Retr.*, vol. 8, no. 1, pp. 1–125, 2014, doi: 10.1561/1500000033.

[7] K. Karako, Y. Chen, P. Song, and W. Tang, "Super-aged society: Constructing an integrated information platform of self-recording lifelogs and medical records to support health care in Japan," *BioSci. Trends*, vol. 13, no. 3, pp. 276–278, Jun. 2019, doi: 10.5582/bst.2019.01124.

[8] G. Meditskos, P.-M. Plans, T. G. Stavropoulos, J. Benois-Pineau, V. Buso, and I. Kompatsiaris, "Multi-modal activity recognition from egocentric vision, semantic enrichment and lifelogging applications for the care of dementia," *J. Vis. Commun. Image Represent.*, vol. 51, pp. 169–190, Feb. 2018, doi: 10.1016/j.jvcir.2018.01.009.

[9] O. Gelonch, M. Ribera, N. Codern-Bové, S. Ramos, M. Quintana, G. Chico, N. Cerulla, P. Lafarga, P. Radeva, and M. Garolera, "Acceptability of a lifelogging wearable camera in older adults with mild cognitive impairment: A mixed-method study," *BMC Geriatrics*, vol. 19, no. 1, pp. 1–10, Dec. 2019, doi: 10.1186/s12877-019-1132-0.

[10] K. Aizawa, "Multimedia FoodLog: Diverse applications from self-monitoring to social contributions," *ITE Trans. Media Technol. Appl.*, vol. 1, no. 3, pp. 214–219, 2013, doi: 10.3169/mta.1.214.

[11] K. G. Stanley and N. D. Osgood, "The potential of sensor-based monitoring as a tool for health care, health promotion, and research," *Ann. Family Med.*, vol. 9, no. 4, pp. 296–298, Jul. 2011, doi: 10.1370/afm.1292.

[12] K. Aizawa, Y. Maruyama, H. Li, and C. Morikawa, "Food balance estimation by using personal dietary tendencies in a multimedia food log," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 2176–2185, Dec. 2013, doi: 10.1109/TMM.2013.2271474.

[13] Y. Kong and Y. Fu, "Human action recognition and prediction: A survey," 2018, *arXiv:1806.11230*. [Online]. Available: http://arxiv.org/abs/1806.11230

[14] J. Yang, J. Lee, and J. Choi, "Activity recognition based on RFID object usage for smart mobile devices," *J. Comput. Sci. Technol.*, vol. 26, no. 2, pp. 239–246, 2011.

[15] Y. Liang, X. Zhou, Z. Yu, and B. Guo, "Energy-efficient motion related activity recognition on mobile devices for pervasive healthcare," *Mobile Netw. Appl.*, vol. 19, no. 3, pp. 303–317, Jun. 2014.

[16] C. F. Crispim-Junior, V. Buso, K. Avgerinakis, G. Meditskos, A. Briassouli, J. Benois-Pineau, I. Y. Kompatsiaris, and F. Bremond, "Semantic event fusion of different visual modality concepts for activity recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1598–1611, Aug. 2016.

[17] Z. Hussain, M. Sheng, and W. E. Zhang, "Different approaches for human activity recognition: A survey," 2019, *arXiv:1906.05074*. [Online]. Available: http://arxiv.org/abs/1906.05074

[18] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 6, pp. 790–808, Nov. 2012.

[19] T.-H.-C. Nguyen, J.-C. Nebel, and F. Flórez-Revuelta, "Recognition of activities of daily living with egocentric vision: A review," *Sensors*, vol. 16, no. 1, p. 72, Jan. 2016, doi: 10.3390/s16010072.

[20] M. Ramanathan, W.-Y. Yau, and E. K. Teoh, "Human action recognition with video data: Research and evaluation challenges," *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 5, pp. 650–663, Oct. 2014.

[21] R. Poppe, "A survey on vision-based human action recognition," *Image Vis. Comput.*, vol. 28, no. 6, pp. 976–990, Jun. 2010.

[22] Y. Kong, Z. Tao, and Y. Fu, "Deep sequential context networks for action prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 3662–3670, doi: 10.1109/CVPR.2017.390.

[23] Q. Shi, L. Cheng, L. Wang, and A. Smola, "Human action segmentation and recognition using discriminative semi-Markov models," *Int. J. Comput. Vis.*, vol. 93, no. 1, pp. 22–32, May 2011.

[24] C. Feichtenhofer, A. Pinz, and R. P. Wildes, "Spatiotemporal multiplier networks for video action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4768–4777.

[25] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 3, pp. 257–267, Mar. 2001.

[26] D. Weinland, R. Ronfard, and E. Boyer, "Free viewpoint action recognition using motion history volumes," *Comput. Vis. Image Understand.*, vol. 104, nos. 2–3, pp. 249–257, Nov. 2006.

[27] S. Kumari and S. K. Mitra, "Human action recognition using DFT," in *Proc. 3rd Nat. Conf. Comput. Vis., Pattern Recognit., Image Process. Graph.*, Dec. 2011, pp. 239–242.

[28] T. Ahmad, J. Rafique, H. Muazzam, and T. Rizvi, "Using discrete cosine transform based features for human action recognition," *J. Image Graph.*, vol. 3, no. 2, pp. 96–101, 2015.

[29] X. Peng, L. Wang, X. Wang, and Y. Qiao, "Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice," 2014, *arXiv:1405.4506*. [Online]. Available: http://arxiv.org/abs/1405.4506

[30] D. D. Dawn and S. H. Shaikh, "A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector," *Vis. Comput.*, vol. 32, no. 3, pp. 289–306, Mar. 2016.

[31] H. Wang and C. Schmid, "Action recognition with improved trajectories," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 3551–3558, doi: 10.1109/ICCV.2013.441.

[32] X. Yang and Y. Tian, "Super normal vector for activity recognition using depth sequences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 804–811, doi: 10.1109/CVPR.2014.108.

[33] M. Li, H. Leung, and H. P. H. Shum, "Human action recognition via skeletal and depth based feature fusion," in *Proc. 9th Int. Conf. Motion Games*, Oct. 2016, pp. 123–132.

[34] A. Jalal, S. Kamal, and D. Kim, "A depth video-based human detection and activity recognition using multi-features and embedded hidden Markov models for health care monitoring systems," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 4, no. 4, p. 54, 2017.

[35] A. Farooq, F. Farooq, and A. V. Le, "Human action recognition via depth maps body parts of action," *TIIS*, vol. 12, no. 5, pp. 2327–2347, 2018.

[36] A. Kamel, B. Sheng, P. Yang, P. Li, R. Shen, and D. D. Feng, "Deep convolutional neural networks for human action recognition using depth maps and postures," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 9, pp. 1806–1819, Sep. 2019.

[37] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, "Learning realistic human actions from movies," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[38] P. Natarajan and R. Nevatia, "Online, real-time tracking and recognition of human actions," in *Proc. IEEE Workshop Motion Video Comput.*, Jan. 2008, pp. 1–8.

[39] H.-I. Suk, B.-K. Sin, and S.-W. Lee, "Hand gesture recognition based on dynamic Bayesian network framework," *Pattern Recognit.*, vol. 43, no. 9, pp. 3059–3072, Sep. 2010.

[40] L. Xu, W. Yang, Y. Cao, and Q. Li, "Human activity recognition based on random forests," in *Proc. 13th Int. Conf. Natural Comput., Fuzzy Syst. Knowl. Discovery (ICNC-FSKD)*, Jul. 2017, pp. 548–553.

[41] J. C. Núñez, R. Cabido, J. J. Pantrigo, A. S. Montemayor, and J. F. Vélez, "Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition," *Pattern Recognit.*, vol. 76, pp. 80–94, Apr. 2018.

[42] Y. Lu, Y. Wei, L. Liu, J. Zhong, L. Sun, and Y. Liu, "Towards unsupervised physical activity recognition using smartphone accelerometers," *Multimedia Tools Appl.*, vol. 76, no. 8, pp. 10701–10719, Apr. 2017.

[43] A. Jordao, A. C. Nazare, Jr., J. Sena, and W. R. Schwartz, "Human activity recognition based on wearable sensor data: A standardization of the state-of-the-art," 2018, *arXiv:1806.05226*. [Online]. Available: http://arxiv.org/abs/1806.05226

[44] I. A. Lawal, F. Poiesi, D. Anguita, and A. Cavallaro, "Support vector motion clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 11, pp. 2395–2408, Nov. 2017.

[45] O. S. Eyobu and D. Han, "Feature representation and data augmentation for human activity classification based on wearable IMU sensor data using a deep LSTM neural network," *Sensors*, vol. 18, no. 9, p. 2892, Aug. 2018.

[46] T. Sztyler, H. Stuckenschmidt, and W. Petrich, "Position-aware activity recognition with wearable devices," *Pervasive Mobile Comput.*, vol. 38, pp. 281–295, Jul. 2017.

[47] I. A. Lawal and S. Bano, "Deep human activity recognition using wearable sensors," in *Proc. 12th ACM Int. Conf. Pervasive Technol. Rel. Assistive Environ.*, 2019, pp. 45–48.

[48] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Human activity recognition with inertial sensors using a deep learning approach," in *Proc. IEEE SENSORS*, Oct. 2016, pp. 1–3.

[49] G. Meditskos, P.-M. Plans, T. G. Stavropoulos, J. Benois-Pineau, V. Buso, and I. Kompatsiaris, "Multi-modal activity recognition from egocentric vision, semantic enrichment and lifelogging applications for the care of dementia," *J. Vis. Commun. Image Represent.*, vol. 51, pp. 169–190, Feb. 2018.

[50] E. Zdravevski, B. R. Stojkoska, M. Standl, and H. Schulz, "Automatic machine-learning based identification of jogging periods from accelerometer measurements of adolescents under field conditions," *PLoS ONE*, vol. 12, no. 9, Sep. 2017, Art. no. e0184216.

[51] K. Ozcan and S. Velipasalar, "Wearable camera- and accelerometer-based fall detection on portable devices," *IEEE Embedded Syst. Lett.*, vol. 8, no. 1, pp. 6–9, Mar. 2016.

[52] S. Song, V. Chandrasekhar, B. Mandal, L. Li, J.-H. Lim, G. S. Babu, P. P. San, and N.-M. Cheung, "Multimodal multi-stream deep learning for egocentric activity recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2016, pp. 24–31.

[53] H. F. Nweke, Y. W. Teh, G. Mujtaba, and M. A. Al-garadi, "Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions," *Inf. Fusion*, vol. 46, pp. 147–170, Mar. 2019.

[54] B. Cvetković, R. Szeklicki, V. Janko, P. Lutomski, and M. Luštrek, "Real-time activity monitoring with a wristband and a smartphone," *Inf. Fusion*, vol. 43, pp. 77–93, Sep. 2018.

[55] C. Gurrin, H. Joho, F. Hopfgartner, L. Zhou, V.-T. Ninh, T.-K. Le, R. Al-Batal, D.-T. Dang-Nguyen, and G. Healy, "Advances in lifelog data organisation and retrieval at the NTCIR-14 Lifelog-3 task," in *Proc. 14th Int. Conf. NII Testbeds Community Inf. Access Res. (NTCIR)*, vol. 11966. Tokyo, Japan: Springer, Nov. 2019, pp. 16–28. [Online]. Available: http://eprints.whiterose.ac.uk/152522/

[56] D.-T. Dang-Nguyen, L. Piras, M. Riegler, L. Zhou, M. Lux, M.-T. Tran, T.-K. Le, V.-T. Ninh, and C. Gurrin, "Overview of ImageCLEFlifelog 2019: Solve my life puzzle and lifelog moment retrieval," in *Proc. Conf. Labs Eval. Forum, CEUR-Workshop (CLEF)*, vol. 2380, 2019, pp. 9–12.

[57] T.-K. Le, V.-T. Ninh, D.-T. Dang-Nguyen, M.-T. Tran, L. Zhou, P. Redondo, S. Smyth, and C. Gurrin, "Lifeseeker: Interactive lifelog search engine at LSC 2019," in *Proc. ACM Workshop Lifelog Search Challenge (LSC)*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 37–40, doi: 10.1145/3326460.3329162.

[58] C. Chang, M. Fu, H. Huang, and H. Chen, "An interactive approach to integrating external textual knowledge for multimodal lifelog retrieval," in *Proc. ACM Workshop Lifelog Search Challenge (LSC, ICMR)*, C. Gurrin, K. Schöffmann, H. Joho, D. Dang-Nguyen, M. Riegler, and L. Piras, Eds., Ottawa, ON, Canada, Jun. 2019, pp. 41–44, doi: 10.1145/3326460.3329163.

[59] S. Chowdhury, P. J. McParlane, M. S. Ferdous, and J. Jose, "'My day in review': Visually summarising noisy lifelog data," in *Proc. 5th ACM Int. Conf. Multimedia Retr. (ICMR)*. New York, NY, USA: Association for Computing Machinery, 2015, pp. 607–610, doi: 10.1145/2671188.2749393.

[60] M. Tournadre, G. Dupont, V. Pauwels, B. C. M. Lmami, and A. Gînsca, "A multimedia modular approach to lifelog moment retrieval," in *Proc. Working Notes CLEF, Conf. Labs Eval. Forum, CEUR Workshop*, vol. 2380, L. Cappellato, N. Ferro, D. E. Losada, and H. Müller, Eds., Lugano, Switzerland, 2019, pp. 1–13.

[61] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," *arXiv:1808.01974*. [Online]. Available: http://arxiv.org/abs/1808.01974

[62] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 1. Cambridge, MA, USA: MIT Press, 2014, pp. 487–495.

[63] R. Ribeiro, A. J. R. Neves, and J. L. Oliveira, "UA.PT bioinformatics at ImageCLEF 2019: Lifelog moment retrieval based on image annotation and natural language processing," 2019.

[64] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[65] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1452–1464, Jun. 2018.

[66] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft COCO: Common objects in context," 2014, *arXiv:1405.0312*. [Online]. Available: http://arxiv.org/abs/1405.0312

[67] I. Krasin, T. Duerig, N. Alldrin, A. Veit, S. Abu-El-Haija, S. Belongie, D. Cai, Z. Feng, V. Ferrari, and V. Gomes, "OpenImages: A public dataset for large-scale multi-label and multi-class image classification," 2016.

[68] G. Patterson, C. Xu, H. Su, and J. Hays, "The SUN attribute database: Beyond categories for deeper scene understanding," *Int. J. Comput. Vis.*, vol. 108, nos. 1–2, pp. 59–81, May 2014.

[69] F. Abdallah, G. Feki, B. A. Anis, and C. B. Amar, "Big data for lifelog moments retrieval improvement," 2019.

[70] J. Lin and J. H. Lim, "VCI2R at the NTCIR-13 lifelog-2 lifelog semantic access task," in *Proc. 13th NTCIR Conf. Eval. Inf. Access Technol.*, 2017, pp. 1–5.

[71] L. Xia, Y. Ma, and W. Fan, "VTIR at the NTCIR-12 2016 lifelog semantic access task," in *Proc. NTCIR*, 2016, pp. 1–4.

[72] S. Yamamoto, T. Nishimura, Y. Takimoto, T. Inoue, and H. Toda, "PBG at the NTCIR-13 lifelog-2 LAT, LSAT , and LEST tasks," in *Proc. 13th NTCIR Conf. Eval. Inf. Access Technol.*, 2017, pp. 1–8.

[73] L. Zhou, L. Piras, M. Riegler, G. Boato, D.-T. Dang-Nguyen, and C. Gurrin, "Organizer team at ImageCLEFlifelog 2017: Baseline approaches for lifelog retrieval and summarization," in *Proc. CLEF*, 2017, pp. 1–11.

[74] P. Zhou, C. Bai, and J. Xia, "ZJUTCVR team at ImageCLEFlifelog2019 lifelog moment retrieval task," in *Proc. CLEF*, 2019, pp. 1–11.

[75] S. Taubert, S. Kahl, D. Kowerko, and M. Eibl, "Automated lifelog moment retrieval based on image segmentation and similarity scores," in *Proc. CLEF*, 2019.

[76] T. Suzuki and D. Ikeda, "Smart lifelog retrieval system with habit-based concepts and moment visualization," in *Proc. 14th NTCIR Conf.*, 2019, pp. 1–8.

[77] Y. J. Lee, J. Ghosh, and K. Grauman, "Discovering important people and objects for egocentric video summarization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1346–1353.

[78] Z. Guo, L. Gao, X. Zhen, F. Zou, F. Shen, and K. Zheng, "Spatial and temporal scoring for egocentric video summarization," *Neurocomputing*, vol. 208, pp. 299–308, Oct. 2016, doi: 10.1016/j.neucom.2016.03.083.

[79] Z. Lu and K. Grauman, "Story-driven summarization for egocentric video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2714–2721.

[80] M. Sun, A. Farhadi, B. Taskar, and S. Seitz, "Summarizing unconstrained videos using salient montages," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2256–2269, Nov. 2017.

[81] S. Jadon and M. Jasim, "Video summarization using keyframe extraction and video skimming," EasyChair, Tech. Rep. 1181.

[82] A. Sahu and A. S. Chowdhury, "Summarizing egocentric videos using deep features and optimal clustering," *Neurocomputing*, vol. 398, pp. 209–221, Jul. 2020.

[83] B. Mahasseni, M. Lam, and S. Todorovic, "Unsupervised video summarization with adversarial LSTM networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2982–2991.

[84] Z. Gao, G. Lu, and P. Yan, "Key-frame selection for video summarization: An approach of multidimensional time series analysis," *Multidimensional Syst. Signal Process.*, vol. 29, no. 4, pp. 1485–1505, Oct. 2018.

[85] T. Jacquemard, P. Novitzky, F. O'Brolcháin, A. F. Smeaton, and B. Gordijn, "Challenges and opportunities of lifelog technologies: A literature review and critical analysis," *Sci. Eng. Ethics*, vol. 20, no. 2, pp. 379–409, Jun. 2014.

[86] R. Rawassizadeh and A. M. Tjoa, "Securing shareable life-logs," in *Proc. IEEE 2nd Int. Conf. Social Comput.*, Aug. 2010, pp. 1105–1110.

[87] I. Askoxylakis, I. Brown, P. Dickman, M. Friedewald, K. Irion, E. Kosta, M. Langheinrich, P. McCarthy, D. Osimo, and S. Papiotis, "To log or not to log? Risks and benefits of emerging life-logging applications," in *Proc. ENISA*.

[88] T. D. *et al.*, "Looking into the crystallball: A report on emerging technologies and security challenges," TENSOR.

[89] N. E. Petroulakis, I. G. Askoxylakis, and T. Tryfonas, "Life-logging in smart environments: Challenges and security threats," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 5680–5684.

[90] A. Allen, "Dredging up the past: Lifelogging, memory and surveillance," Fac. Scholarship Penn Law, Tech. Rep. 75.

AMEL KSIBI received the B.S., M.S., and Ph.D. degrees in computer engineering from the National School of Engineering of Sfax (ENIS), Sfax University, Tunisia, in 2008, 2010, and 2014, respectively. She spent three years at ENIS as a Teaching Assistant before joining the Higher Institute of Computer Science and Multimedia Gabes (ISIMG) as a Permanent Assistant, in 2013. She joined the Computer Science Department, Umm Qura University (UQU), as an Assistant Professor, in 2014. She joined Princess Nourah Bint Abdulrahman University, in 2018, where she is currently an Assistant Professor with the Department of Information Systems, College of Computer Sciences and Information. Her research interests include computer vision and image and video analysis. These research activities are centered on deep learning applied to lifelog image and video understanding, indexing and retrieval.

ALA SALEH D. ALLUHAIDAN (Member, IEEE) received the B.Sc. degree in computer science from Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia, the M.Sc. degree in computer information systems from Grand Valley State University, Allendale, MI USA, and the Ph.D. degree in information systems and technology from Claremont Graduate University, Claremont, CA, USA. She is currently an Assistant Professor with the Department of Information Systems, Princess Nourah Bint Abdulrahman University. Her research interests include health informatics, big data analytics, and machine learning.

AMINA SALHI received the B.S. and M.S. degrees in computer sciences from the University of Guelma, Algeria, in 2008 and 2010, respectively, and the Ph.D. degree in image processing and computer vision from the University of BMA, Algeria, in 2017. Since 2018, she has been an Assistant Professor with Princess Norah Bint Abdulrahman University (PNU), Riyadh, Saudi Arabia. Her research interests include computer vision, image processing, and biometry.

SAHAR A. EL-RAHMAN (Senior Member, IEEE) received the M.Sc. degree in an AI technique applied to machine aided translation and the Ph.D. degree in reconstruction of high-resolution image from a set of low-resolution images from the Faculty of Engineering-Shoubra, Benha University, Cairo, Egypt, in 2003 and 2008, respectively. Since 2008, she has been an Assistant Professor with the Faculty of Engineering-Shoubra, Benha University. She is currently an Assistant Professor with the College of Computer and Information System, Princess Nourah Bint Abdulrahman University, Saudi Arabia. Her research interests include computer vision, image processing, information security, human–computer interaction, e-health, big data, and cloud computing.

• • •