Proceedings of the Pakistan Academy of Sciences:
A. Physical and Computational Sciences 56 (3): 1-16 (2019)

ISSN: 2518-4245 (print), 2518-4253 (online)

Pakistan Academy of Sciences

Review Article

An Insight of Smartphone-Based Lifelogging Research: Issues, Challenges, and Research Opportunities

Shaukat Ali[1*], Shah Khusro[1], Akif Khan[1], Inayat Khan[1], and Salman Faiz Solheria[2]

[1]Department of Computer Science, University of Peshawar, Pakistan
[2]Sarhad University of Science & Technology (SUIT), Peshawar Pakistan

Abstract: The technological advancements have turned smartphone into a de-facto lifelogging device and fostered smartphone-based lifelogging (SBL) research as a mainstream activity. A SBL system can capture and store information about peoples' daily life activities, behaviours, interactions, contexts, etc., into a comprehensive personal lifelogs. The smartphone-based personal lifelogs can be of interest to the different stakeholders including users, information sciences researchers, and policies/decisions makers in governments and organizations because of the availability of information for solving different real-world problems (e.g., memory augmentation, medicine and health care, business and commerce, government policy making, society development, etc.), which would otherwise be impossible. Despite of potential advantages, SBL research is in its early stages and has several issues and challenges, which impedes its large-scale adaptation. Therefore, more research efforts are needed in the SBL domain. This paper is aimed to provide an insight review of SBL research and highlight several issues, challenges, and research opportunities. Firstly, the smartphone is demonstrated as a de-facto lifelogging device and available SBL research is analysed to highlight research shortcomings and gaps. Secondly, a generalized architecture for SBL systems is presented to unify and accelerate the research and development efforts. Lastly, several issues and challenges are highlighted, which could be turned into potential research opportunities. This paper can be a valuable resource for new researchers to find research topics to contribute and explore the SBL research area.

Keywords: Smartphone, Lifelogging, Sensors, Privacy & Security, Semantic Organization, Memory Augmentation.

1. INTRODUCTION

Lifelogging - a step towards "Memories for Life" grand challenge for computing research [1] - refer to using of computer technology for comprehensively digitally archiving of peoples' life time experiences in multimedia format for a variety of use-case including enabling people to mine and infer knowledge about their livings [2-4]. The most appropriate definition of lifelogging is "a form of pervasive computing, consisting of a unified digital record of the totality of an individual's experiences, captured multi-modally through digital sensors and stored permanently as a personal multimedia archive" [2]. A lifelogging system has to operate both continuously and passively to automatically capture both content and contextual information without requiring any explicit users' efforts or interventions [5] from different information sources including web pages browsed, sent and received emails and SMSs, electronic calendar entries, dialed and received phone calls, downloaded and listened audios and videos, photographs, contextual and environmental information captured via sensors (e.g., locations using GPS, etc.), etc. Lifelogging can be either total capture or selective capture [6]. In any type, lifelogging is aimed to resolve challenges of enhancing peoples' performances by providing complementary digital assistance.

The lifelogging systems are mostly developed using wearable computing technologies, relying on external capturing and sensing devices to automatically record peoples' daily life activities and contextual information [7-9]. However, the wearable lifelogging systems are providing

Received: June 2019; Accepted: September 2019
*Corresponding Author: Shaukat Ali <shoonikhan@uop.edu.pk>

limited features with additional disadvantages of overloading users with extra devices, which could overload users, produce hurdles and troubles in performing daily life activities, and could be sources of social critiques. The advent of ubiquitous computing technologies has shown that verbatim capturing and storing of one's total life experiences is possible and affordable to provide a digital memory [10]. Using ubiquitous computing devices for lifelogging is necessary because occurrences of significant events/activities are independent of a particular schedule and location. Smartphone (SP) is proven highly ubiquitous computing device and is becoming commonplace by showing its presence in the pocket of almost every individual today [11]. They are our constant companions and know us very well beyond our imagination by collecting a wide variety of effective information about almost all aspects of our lives as well as contextual and environmental information [11, 12]. The SP's portability, built-in features, and capability of integration with other technologies make it not hard to believe SP as a de-facto lifelogging device [12, 13]. A SP integrates enormous computing technologies especially sensors, storage, and networking. The sensory capabilities enable SP to continuously and unobtrusively capture user content and contextual information related to our daily life activities, actions, and environments [14, 15]. The storage technology enables SP to store a large amount of data for a long time. The networking technologies enable SP to connect to a vast computing and storage media (e.g., cloud) for data processing and storage [13].

To date, we have a few SBL researches from academia and organizations [13, 16, 17]. The SBL research is majorly focused on developing applications, emphasizing on exploiting SP capabilities to fulfill objectives of lifelogging. However, the technological advancements in SP have made the scope of SBL wider than traditional lifelogging technologies and could be a step towards capturing and recording of voluminous totality of life experiences information into a lifelog. The SBL applications can be advantageous in several ways including bringing changes in behaviors of people by analyzing past behaviors, finding peoples' preferences for dinning in a region, enabling people to remember names of previously visited places, etc. However, SP to fulfill design requirements of lifelogging systems (i.e., seamless incorporation in daily life, resources efficiency, security, long-term lifelog data preservation, and retrieving information from lifelog), a number of issues and challenges are needed to be solved [13]. These are mainly regarding capturing, processing, storing, index, retrieval and usage, visualization, ownership, ethics, etc., of huge lifelogs, which needs to be continually addressed by the researchers with the advancements in sensors, computer science, and information management technologies [18].

The objective of the paper is to review and analyze the SBL research and presenting in a chronological order to highlight research gaps, issues and challenges, and research opportunities. To the best of our knowledge, this paper is the first attempt of providing a detailed review of SBL research. However, the main contributions of the paper are: (1) providing detailed insight analysis of the SBL research using their methodologies, functionalities/features, lifelog information capturing for harnessing a personal lifelog, and retrieval; (2) classifying SBL research into novel taxonomies to identify research gaps; (3) proposing of novel generic architecture for SBL systems to unify research efforts and increase SBL systems development trend; (4) highlighting a number of issues and challenges in the SBL research, which can be turned into potential research opportunities to help new researchers in finding research topics.

2. SMARTPHONE AND LIFELOGGING

The SP is a technologically advanced type of mobile phones that accumulates the features of mobile phone (i.e., voice calls, and SMS) and Personal Digital Assistance (PDA) (i.e., office management, web browsing, email, etc.) [15, 18, 19]. SPs can be a vision of Memex and step towards capturing "totality of life experiences" by offering novel opportunities to ubiquitously and unobtrusively record nearly all aspects of a person's daily life activities and events in a verbatim and unbiased way to construct and preserve a long-term digital memory [10, 18]. The SPs are tremendously improved technologically in the past few years. The SPs weight is reduced (e.g., Apple iPhone 11 is of 194g) to improve mobility and portability while the display size is increased for improved visualization. The battery power is increased to 5000mAh (e.g., Samsung Galaxy M30) to meet power needs of SP's applications and platforms. The increased battery power enables developers to develop applications to exploit full potentials of a SP to solve complex real-world problems. The SPs provide easier and advanced GUIs that are fully multimedia loaded, easy to navigate, easy to understand, touch enabled virtual keyboard, and user-friendly. The multi-touch technology is available in SPs to track more than one touches at the same time [20]. The SP processor power is increased and octa-core processors are available in SPs (e.g., Google Pixel 3A) to execute complex applications and process large data. Similarly, SPs are integrated with Graphical Processing Unit (GPU) to execute graphical calculations and transformations, and reduces the burden of Central Processing Unit (CPU) for enhanced performance. The SPs have internal data-storage capacity of 512 GB (e.g., Samsung Galaxy Note 10+), which can log enormous amount of data about a user's activities, contacts, calendar data, etc. For example, a SP with 128 GB storage can store images for more than 3 years if taken with a frequency of 1.65 million images per year [19]. The continued advancements and miniaturization in storage technologies have enabled the development of slim, lightweight, and high volume removable storage metaphors (e.g., microSD, microSDCX, and small form-factor disks) for the SPs. As it was predicted in 2006 [21], the removable storage capacity has reached up-to 2TB in a single card, which is large enough to store digital information of a person's life [22].

The advancements in networking capability have enabled SPs to access and exploit the free storage and processing capabilities offered by various cloud services for transferring, storing, and processing excessively a variety of personal data to create digital memories [12, 13]. In addition, SPs can support RAM up-to 12 GBs (e.g., Samsung Galaxy Note 10+). The high RAM capacity not only contributes in enhancing speed of a SP but also enables execution of complex applications including big data analytics. The SPs are embedded with a significant set of sensors (e.g., accelerometer, gyroscope, proximity, compass, barometer, fingerprint, etc.) with the same power and sophistications of external sensors and the number of sensors is expected to increase with the passage of time [15]. The integration of sensors turns a SP into a life-centric sensor to capture a variety of contextual and environmental information to effectively depict peoples' daily life activities and events such as where we go, what we do, who we meet and communicate, what information we use, etc. The SP's sensors have several advantages over wearable sensors [15, 23] and could have potential applications for capturing lifelog information [18]. A detailed discussion of SP's technological developments, sensors available in SPs and their capabilities to capture lifelog and contextual information, conformance of SP as a de-facto lifelogging device by comparison with dedicated wearable lifelogging devices (e.g., SenseCam [7]) can be found in our previous research work [18].

Realizing the technological advancements of SPs and its advantageous characteristics over dedicated wearable lifelogging devices, a few researchers have presented SP-based lifelogging systems by either exploiting the entire set of capabilities (i.e., sensing, processing, storing, and networking) of SPs or using SPs in conjunction with wearable devices or remote computing infrastructures for monitoring and storing lifestyles and activities [12, 24]. The SP users have already witnessed the first generation of SP lifelogging apps (e.g., Saga, Moves, and Rove etc.) for passive capturing and archiving of specific types of life experiences and contextual and environmental data. For example, Saga and Move passively captures and generates lifelog of users' daily life activities data only but do not capture visual contents. In addition, several apps (e.g., Instant, Loca, Fit time, Sleepy, RescueTime, etc.) are mainly logging users' daily life information about fitness, locations and places visited, sleep duration, weather, tracking programs usage on device. However, in the published researches, Nokia Lifeblog project [16] is the first among the SBL systems. The Nokia Lifeblog provided inspiration for many of the subsequent research efforts in both industry and academia that resulted into the emergence of a new breed of SBL systems with increasing sensing and logging functionalities such as Pensevie [25], MyExperience [26], Experience Explorer [11], UbiqLog [13], etc. A satellite view comparison of the SBL researches/systems is shown in Table 1. The researchers agree on the voluminous capturing and long-term storing of lifelog information. However, they have shortcoming of using a specific set of sensors to capture specific and limited lifelog items, storing lifelog dataset on remote machines in fixed schemas, domain specific retrieval and visualizations, etc., which impedes the potentialities and advantages of SBL.

3. TAXONOMIES OF SMARTPHONE-BASED LIFELOGGING RESEARCH

This section analyzes the on-hand SBL research and presents detailed taxonomies. The taxonomies are proposing categorization methods for SBL research using various attributes/aspects such as basic architecture, role of SP, scope of lifelogging, sensor placement and sensing mechanism, etc. The taxonomies are aimed to provide a complete reflection of the present SBL research and the possible future prospects. The proposed taxonomies could also be helpful to evaluate and identify research gaps in the available SBL research. In the taxonomies figures, the solutions reported by researchers are represented with rectangles and the unreported solutions are represented with rounded rectangles, which could be potentially new areas for the future research.

Using sensors deployment and role of SP, the available systems can be categorized roughly

Table 1. Satellite view comparison of the smartphone-based lifelogging systems.

Publication	Sensors	Contextual Information	Lifelog Items	Annotations Technique	Storage	Sharing	Retrieval
Experience Explorer [11]	GPS, WiFi, GSM, Bluetooth, Camera	Location, Time, Neighborhood, Keywords	Pictures	Semi-Automatic	Database - MySQL	Yes (Flicker Photo Service)	PC – Timeline
Nokia Lifeblog [16]	GPS, Camera, metadata	Time, Location, Object Name, Phone Number	Pictures, videos, SMS, MMS, notes, blogs	Automatic	SQLite	Yes (Typepad)	Mobile Phone and PC – Timeline
Memory Book [27]	GPS, Bluetooth, Camera, metadata	Location, Time, Neighborhood	Pictures, text	Automatic	RDF	Not Available	WWW – Timeline
Pensieve [25]	Camera, Microphone	Location, Time, metadata	Pictures, Audios	Semi-Automatic	Lucene Indexer	Yes	PC - Web UI
iRemember [40]	Microphone	Time	Audios	Automatic	Not Available	No	PC – Timeline
Mobile Lifelogger [10]	Accelerometer, GPS, Camera, Microphone, WiFi, Rotation	Location, Time	Pictures, Audios, Activities	Semi-Automatic	Not Available	No	WWW-Timeline
UbiqLog[13]	Not Available	Location, Time	Applications, SMSs, Pictures, Calls	Semi-Automatic	Not Available	Yes	Smartphone-Timeline
SenseSeer [12]	Not Available	Not Available	Not Available	Not Available	Cloud	Yes (Cloud)	Smartphone, WWW
Digital Diary [17]	GPS, Camera, Infrared	Location, Neighborhood	Pictures, Audios	Automatic	SQLite	Yes (cloud)	Smartphone-Timeline
SoundBlogs [9]	Microphone, GPS	Location, Time	Audios	Automatic	Not Available	Yes	Smartphone

into in-situ systems and wearable systems (shown in Fig. 1). Using of wearable technologies has been the prime target of researchers; however, researchers have demonstrated using of SP as alternative to the available wearable technologies [24]. This is usually done by harnessing a weaver with a SP [10]. A SP either alone could be used as a wearable device or could be used in conjunction with other computing devices including wearable sensing devices [11] or body-mounted computers/ laptops [27] for resources intensive processing, and retrieval. Researchers have reported using SP alone as wearable by mounting it on upper part of body (i.e., worn in a helmet on the head [10] or in a lanyard round the neck [24]); whereas, other potential placement areas (i.e., placement on waist and lower part of body) needs to be investigated and explored for potential applications and results. The SP has different roles (i.e., sensing, storage, analysis, sharing, and retrieval) while using in combination with other systems/devices. However, this approach could overburden users that could affect their performances and requires them to have explicit devices and communication channels that collectively increase cost and complexity. In-situ lifelogging means lifelogging in an instrumented environment (called smart environment), where capturing of lifelog information is highly dependent on sensors installed in a local infrastructure. However, in-situ SBL systems can either rely fully on SP inertial sensors or sensors already deployed in an environment or any combination of them, with varying roles of SP. This approach would eliminate users from the problems associated with wearable computing devices. However, operations of in-situ SBL systems would be strictly restricted due to dependency on the instrumented environment. In-situ SBL can track peoples' lives in detail but no research attention has been paid to it to-date. Comparatively, all of the SBL systems are developed using wearable technology by requiring users to keep their SPs close to their bodies at a certain position. However, using SP in combination with both (i.e., wearable and in-situ technologies) and placing SP at different body parts can provide effective and voluminous personal dataset opportunity.

Architecturally, the SBL systems can be classified into two broad categories: distributed and integrated (shown in Fig. 2). In distributed approach, SBL system's functionalities are distributed across SP and a remote server or PC. The SP is used as a front-end device for capturing lifelog and contextual data or for low-level data processing and storage, and remote server or PC is used as back-end for resources intensive processing, analysis, indexing, storage, and retrieval of lifelog information such as [10-12, 25]. The data captured by a SP is processed minorly locally and transferred to remote server using Internet or cellular networks or by physically connecting SP to PC [11, 26].

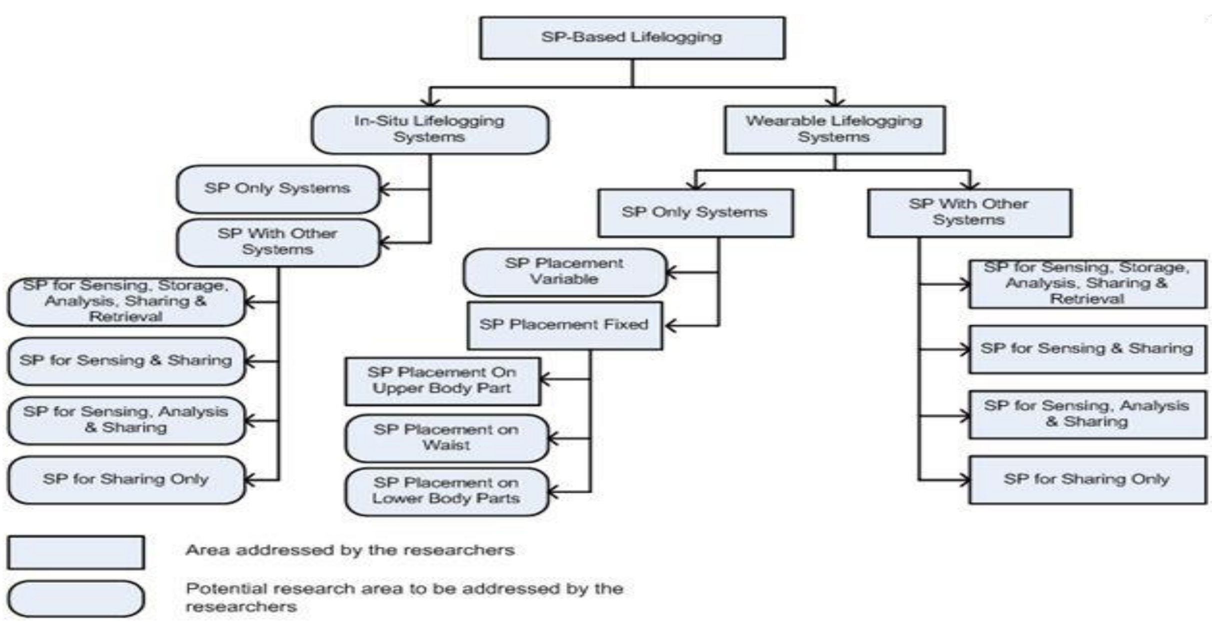

Fig. 1. Taxonomy of smatphone-based lifelogging systems using sensors deployment and role of smatphone in lifelogging.

The distributed approach is the preferred choice of researchers due to resources constraints nature of SP (i.e., low processing power, battery power, storage, etc.) in the past. However, this approach can suffer with a number of problems that can degrade significances of SBL such as (1) transmission delay can be introduced which is not desirable by real-time monitoring systems; (2) data uploading can be problematic in areas where connectivity networks availability cannot be ensured; (3) excessive data transmissions can deplete battery power quickly; and (4) uploading and storing lifelog information remotely can induce privacy and security issues. The integrated approach attempts to exploit the potentials and functionalities of SP for performing all of the lifelogging operations without requiring any external supplements. The integrated approach combines all of the features into a single package and overcomes the problems associated with distributed approach. None of the available SBL systems supports integrated approach. However, UbiqLog [13] and Digital Diary [17] are a bit closer to it. It is not hard to believe that recent technological advancements in SP can attract researchers' interests in developing solutions using integrated approach. However, to fulfill objectives of SBL, its needs to combine both distributed and integrated approaches to provide effective platform for total personal lifelog development and analytics.

Using lifelogging scope, the SBL research can be classified into two broad categories: total capture and selective capture (shown in Fig. 3). Indeed, both of these approaches produce personal lifelogs; however, the size of a lifelog strictly depends on the type of lifelogging. The selective capture is logging of experiences information about specific aspects of a user's life in a few data types (e.g., audios, videos, images, message, notes, etc.) with predefined use-cases. The selective capture is common in SBL due to mining of immediate value from the focused data such as [10, 13, 26]. The total capture is the creation of a unified digital record of multi-modally captured data of totality of life experiences in a variety of data types. The total capture has broad spectrum and can support a number of use-cases. However, the total capture is complex and requires sophistication in gathering, storing, and processing into semantically meaningful and retrieval information for supporting various use-cases. None of the SBL systems supports total capture. However, the recent technological improvements and wide availability of apps for performing daily life activities in SPs can attract researchers' interests for total capturing of life experiences. The total capture can produce comprehensive personal lifelogs and provides an effective environment for analytics to solve a variety of real-world problems in different fields.

Using storage, the SBL research can be classified into two categories: database and ontology (shown in Fig. 4). the database has been the primary choice of SBL researchers for storing lifelogging information since years such as [10, 16, 17]. They have successfully demonstrated storing and retrieval of information from database partly local on SP and majorly on remote backend servers. However, a database has fixed schema and cannot cope with the problem of accommodating new lifelog events and information that could emerge with the changing lifestyles of people over the time [18]. Furthermore, the relational databases are found expensive and technically unwieldy of storing and processing large volume of lifelog information [28]. The set-based retrieval model of

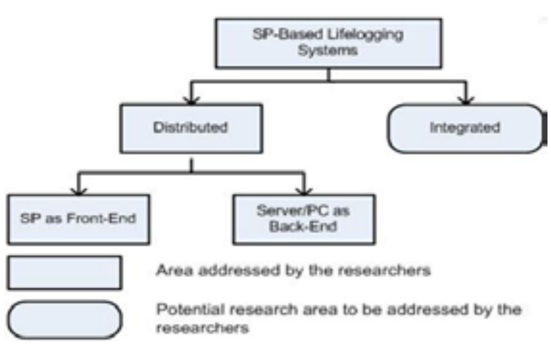

Fig. 2. Taxonomy of smatphone-based lifelogging systems using architecture

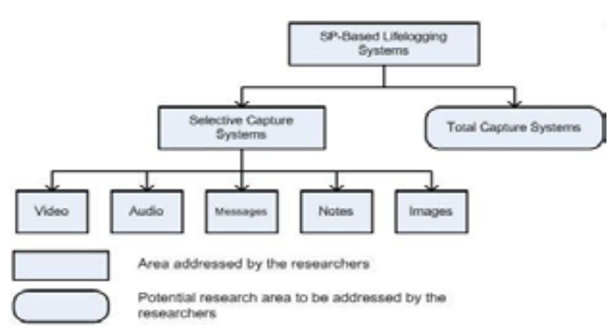

Fig. 3. Taxonomy of smatphone-based lifelogging systems using scope of lifelogging.

relational database quickly becomes unworkable with the increasing size of a lifelog. The huge and growing size of a lifelog from different data sources could make real-world databases susceptible to inconsistency, incompleteness, and noisy data. In addition, daily life information is related with each other in multiple semantic ways in the real-world, which cannot be projected exactly in the relational database technologies. Representing and interlinking lifelog information with the same semantics as they exist in the real world will be helpful in interlinking diverse lifelog information, and developing retrieval models and applications to solve real-world problems [18]. The ontology is a Semantic Web technology that enables to develop a semantically enriched model for lifelog information. The ontology would be more flexible and scalable by adding/modification of new/existing lifelog information and relationships, reasoning and inferencing of new information on the basis of existing information, and covering a wide variety of relationships and data sources [18]. In addition to semantically modeling and organizing lifelog information, the ontology would provide powerful constructs to use and manage personal lifelog information such as querying using SPARQL. A few of researchers have used ontology formalism for lifelog information management in desktop environment such as [29, 30]. The Memory Book [27] has attempted organizing lifelog information in RDF. However, the SBL researchers have not paid attention to using ontology for lifelog information modeling and could be a potential research area for the researchers.

Apart from the above classification taxonomies, the available SBL research can be classified using sharing and retrieval. Lifelog information has two aspects: private and public. The private aspect underpins that personal lifelog information remains in a user ownership and should not shared with others; whereas, public aspect is related with sharing of personal lifelog information using consents of users. Most of the researchers have presented methodologies of sharing lifelog information from SP with their web interfaces such as [10, 27] whereas, a few have postulated sharing on social media such as Flicker [11]. The sharing of personal lifelog information is mandatory to support several effective analytics and applications. However, security and privacy should be ensured while sharing lifelog information. Similarly, the effective SBL practices should provide methods for the implicit swift real-time retrieval of lifelog information to augment human memory and other variety of needs. The effective retrieval requires defining and using of techniques (e.g., data mining, machine learning, visualization approaches, etc.) depending on applications. Most of the researchers have postulated retrieval using contextual information and timeline display either directly on PCs [11] or with the assistance of the Web [11, 25, 27], which is explicit and not in real-time. Some of the researchers have also demonstrated basic retrieval on smartphone [13, 17]. However, the advancements in SP can attract attentions of the researchers to provide advanced retrieval methods to do things like combining, correlating, cross-referencing, leveraging, data mining from heterogeneous sources, learning, and presenting in appropriate and passive manner.

4. GENERIC ARCHITECTURE

The lack of standard guidelines for SBL freed researchers to propose systems using their own experiences and methodologies. The systems are architecturally different from each other. Thus, creating separate islands and is wastage of resources and time. Therefore, we have proposed a generic architecture for SBL systems using lifelogging systems designing principles [4, 31] that will minimize problems, improve systems' interoperability, and support effective lifelog organizations and analytics. Technically, using SP for end-to-end lifelogging is a complex phenomenon and involves many challenges [32]. Therefore, to simplify understanding & development, the architecture is divided into four modules (shown in Fig. 5) [18]. The architecture is practical and is used as baseline for the development of SLOG framework [18].

4.1 Data Collection Engine

Data Collection Engine (DCE) collects lifelog information from a user's personal information (PI) space. A user's PI space is composed of lifelog objects (e.g., pictures, documents, contacts, emails, calls, etc.) related to a user, and contextual and environmental information, which must be made accessible to a SBL system for instrumentation,

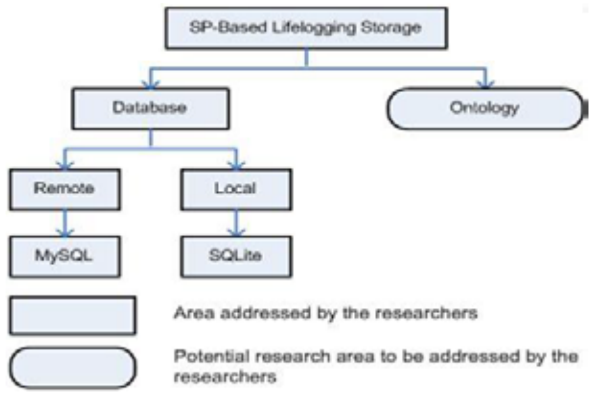

Fig. 4. Taxonomy of smatphone-based lifelogging systems using storage modeling.

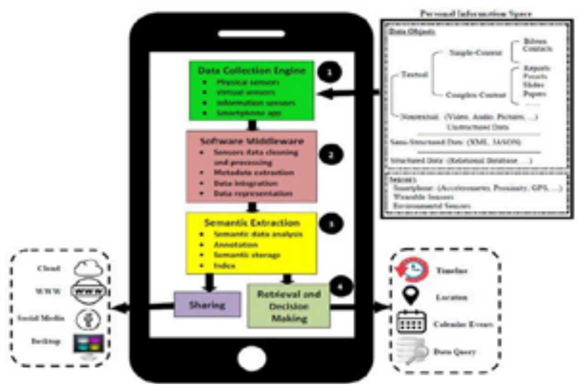

Fig. 5. General architecture for smatphone-based lifelogging systems.

automation, and querying. The DCE can capture lifelog information using different types of sensors (i.e., smartphone, wearable, environmental, etc.) and applications, and relay the captured information for further analysis and storage. However, the number and types of applications and sensors used depends on the type of lifelogging [4, 18]. The DCE would capture information from information sources either reactively on event-based or proactively on polling-based [18]. However, event-based capturing is effective for SP to not waste battery power [18].

4.2 Software Middleware

Software Middleware (SM) would provide components employing different techniques (e.g., machine learning) to preprocess raw sensory data and lifelog items from DCE. For example, aligning data both temporally and spatially, cleansing data from noise, fusion of data into uniform object, computing and utilizing the trust and provenance or reliability of the data streams, transforming unstructured data into a structured format, improving quality of the captured data, transforming the captured data into useful information, and merging/combining captured information into a consistent structure. In addition, contextual information can also be retrieved from smartphone applications such as user's calendar. Similarly, metadata associated with lifelog objects by the parent applications can also provide additional contextual information and is needed to be extracted for enhanced annotations. The lifelog, contextual, and metadata information could be fused and organized in a holistic consistent structure (e.g., JSON) for processing/analysis. Very little research attention has been paid to this area of SBL research to-date. The research experience from the other fields can be used for learning about data quality, trust, and reputation.

4.3 Semantic Extraction and Organization

Semantic Extraction and Organization (SE & O) will provide semantic glue to organize and relate lifelog information with the same semantics as they exist in the real-world and encoded in the human episodic memory [18]. In SBL, to fill the space between raw sensory data and people understandability needs employment of effective semantic extractions, analysis, modeling/ organizing, and reasoning techniques to solve the key challenges of lifelog management and retrieval. The semantic extraction identifies and generates semantic constructs out of lifelog information. The semantic analysis involves several of the structuring, organizing, and summarization processes for mapping lifelog information into more discrete and meaningful discrete units (e.g., events [33, 34]) to determine their uniqueness and regularity within the lifelogger's lifestyle. Semantically enriching lifelog data at the event level involves annotation using contextual and metadata information to describe and relate lifelog information in a more meaningful way that will not only increase data representations and understandability but will also enhance retrieval of specific lifelog information from a lifelog. All of the information and annotations in a lifelog can be indexed for improved retrieval performances. The SE & O is an important activity for SBL to completely exploit the gold ores of lifelogs but has not received significant research attentions to-date.

4.4 Retrieval and Sharing

A SBL lifelog provides primitives of querying implicitly and explicitly numerously (i.e., content-based information or context-based information) for searching, browsing, summarization, and recommendation using a number of information such as time, location, calendar event, and proximity to support a number of use-cases. A lifelog retrieval model can be inspired from the 5Rs of memory access proposed by [6]. In addition, access methodology and HCI factor should be considered while defining a use-case. To fulfill the social aspect of lifelogging, lifelog information can be shared with friends, colleagues, and family members for numerous purposes. The lifelog information can be shared using a number of techniques due to storage limitation of SP for providing global access to personal lifelog and web blog for creating personal digital diaries. The security and privacy should be kept in mind while retrieving and sharing.

5. CHALLENGES AND OPPORTUNITIES

The SBL provides unique opportunities of creating comprehensive personal lifelogs generation. Like any other new technology, SBL can excite and inculcate worries and concerns [32]. One can legitimately ask: why SBL has not made to the mainstream market yet? How knowledge hidden in a lifelog can be valuable for the stockholders? Lack of complete development of the technology is by itself a big challenge, making it difficult to describe clearly the legitimate use of the technology by the stockholders. Functions of the technology are uncertain and not clearly described to-date including controlling functions of the technology and owning personal lifelogs, affects of SBL on an individuals' identities and behaviors and functions, the interoperation between technological memory and biological memory, etc.

In addition, the ubiquitous nature and technological advancements of SPs could make capturing totality of life experiences possible but would intensify the associated challenges, which will make SBL impractical and vulnerable if proper procedures, methods, and policies are not defined. The uncertainties and advancements will pose serious issues and challenges to SBL researchers regarding capturing, storage, searching, analysis, sharing, visualization, and ownership. The challenges are unique from SBL perspectives; therefore, developing solutions for the challenges are vital for the success of SBL. Certainly, the challenges will provide research opportunities for the researchers to contribute into SBL by developing valuable systems.

5.1 Data Capturing and Merging

The SBL systems have either used a single sensor (e.g., microphone [9][40]) or combination of a few sensors (e.g., GPS and camera [16]) for capturing and annotating lifelog information. Technically, using broad range of sensors can provide a large-scale lifelog by capturing information about different aspects of life events. However, fusing either multiple sensors data or sensors data with other information sources is crucially a difficult task. The sensors data fusion requires careful data cleansing, alignment, and temporal normalization [4]. Therefore, knowledge and research efforts from sensors data fusion domain [35, 36] can be used to develop effective schemes for fusing data from various SP sensors and sources to obtain improved information (i.e., less expensive, higher quality, and more relevant) for accurate estimations of contexts and contents. Instead of using a static combination of sensors, adaptive and dynamic selection of sensors and their sampling rates could be employed to produce more energy-efficient solutions. In addition to sensors, relevant information can also be captured from other sources. For example, using phone number to extract information (e.g., name, address, etc.) from a phone contacts and using location coordinates to extract location information (e.g., known name, postal code, etc.) from online repositories. However, specialized techniques are needed to be developed for meaningful information merging from the different and diverse sources into compact information.

5.2 Targeting Fine-Grained Data Events and Activities

The main tasks in a SBL system are determining a set of target events and associating sensory data and other inputs as contexts to the events. The SBL systems have claimed reasonable retrieval rate due to recording a small set of coarse-gained events and activities. However, such approaches would be of little use due to their limited scopes such as automatic diary would be incomplete if it could not recognize meetings in an office environment. Therefore, the legitimate question is what would be the effective set of daily life events and activities that should be recognized and recorded in SBL lifelog for empowering potential applications. There are three potential reasons of defining a potential set of common activities for all of the people. Firstly, human life is dynamic and changes with the passage of time making peoples' interactions and activities different from each other. Secondly, the integration of new technologies in SPs (i.e., sensors and applications) can emerge new opportunities of lifelog information capturing. Finally, SBL can have potential applications of diverse nature. Nevertheless, defining a set of daily life events and activities to be captured could benefit in setting objectives and guidelines for research efforts.

5.3 Data Processing and Storage

The SBL systems are mainly implemented in centralized architecture, where SP is used for data collections and backend servers on the Web or PCs are used for processing and storing of lifelogs. The storage technologies for servers and PCs are improved in storage capacities and I/O speeds. However, centralized architecture will require users to have explicit network connections, devices to store and review them lifelog information. This increases overall cost, security and privacy issues, and bandwidth bottleneck in case of large volume of communication. The performance of data-intensive applications can be increased by optimizing data access techniques including data replications, migration, distribution, and access parallelism [31]. The processing and storage advancements are needed to be extended to SP for implementing distributed data centric storage. This would replicate lifelog on SP and ensure user control, ubiquitous and omnipresent access, and data archiving without relying on additional technologies.

5.4 Data Curation and Organization

Data curation aims at data authentication and quality assurance, archiving, management, long time preservation, and retrieval [31]. The SBL systems emphasize on storing and indexing lifelog information using database technologies. The existing database management tools cannot handle lifelog archive that stores multitude of information which grows exponentially in size and complexity. The limitations of relational databases to store SBL lifelogs are discussed in Section 3. In addition, using relational databases cannot rationalize the sharing aspect of lifelogging due to requiring API access [18]. The NoSQL and Hadoop are non-relational approaches for large and distributed data management and database design for processing and storing voluminous data in parallel across a grid of servers. The advantages of schema-free databases are the empowerment of developers to change the structure of data without rewriting tables and greater flexibility if data is heterogeneously stored. Using relational and schema-free databases can leave semantic gap between the lifelog information stored and their occurrence in the real-world and human memory [18]. In real-world, lifelog information is related with each other in different semantic ways, which could not be exactly represented by these technologies due to their limitations of features and constructs [18]. Therefore, scalable and flexible, high performance, and low-level access storage methods are needed to store lifelogs; containing information different lifelog items and their relationships in a subtle way. The Semantic Web technologies can solve this problem by formulating lifelog information using ontologies and storing them in RDF triple format for advance query, exploration, and connections with other lifelog information at any other place (e.g., LOD) [18].

5.5 Data Analysis and Visualization

The volume of a lifelog is scalable and would grow with the passage of time into a personal big data archive, which will make it difficult for lifelog analysis tasks. Therefore, research contributions from big data analytics can be used in SBL. The researchers have highlighted effective analysis algorithms to deal with the growing volume of a big data and the processor speed is increased following the Moore's Law [31]. From the perspective of big data analytics techniques, increment algorithms have good scalable property; however, not suitable for all machine learning algorithms [31]. Technically, despite of introducing core technology in SP's processors, the processor clock speed is still highly lag behind due to the expected scaling of personal big data archives much faster than processor speed. However, the problem can be handled by the development of parallel computing in SPs. Another solution could be cloud computing, which could combine multiple disparate workloads into a large cluster of processors. The lifelog visualization represents and conveys information hidden in large-scale data set understandably and easily. The SBL systems have data visualization approaches of very low performance in functionalities, scalability, and response time. Therefore, lifelog visualization approaches are needed to be revised to develop new data visualization tools to solve the problems such as Tableau of eBay [31].

5.6 Societal Acceptance

The lifelogging technologies have witnessed negative responses from the society (e.g., Google Glasses [4, 32]). However, SP is a commonplace and can be used by the people in different places and situations openly and freely. Therefore, SBL can have high degree of societal acceptance as compared to its predecessors. However, SBL as a new technology will be acceptable by the society when it finds space in the mainstream. The increasing understandings and highlighting of positive benefits of personal lifelog harnessing, sharing, analytics, and applications can turn peoples' thoughts and can result into its high-level adaptation. The realization and availability of useful lifelog applications could be handy in this regard, which would put SBL into social debates by disclosing its positive benefits.

5.7 Violations of Privacy and Security

The SBL can be tailored with increased privacy implications. Privacy detrimental depends on type of SBL where selective lifelogging would be of less security concern as compared to total lifelogging. Privacy in SBL is intensive and puzzling with no clear definition. Some researchers have regard privacy with control over information in personal lifelog and have developed their own concepts of privacy, and others have aimed of offering recommendations for the developers [4]. The problems become more dangerous due to SP potentially capturing totality of life experiences. Some of privacy concerns could be:

5.7.1 Information Ownership

The important challenges in SBL are ownership and access of personal lifelogs [4, 32, 37]. Solutions to the legitimate questions are needed such as where to store, who owns, what could be the lifetime, who can access, etc. Technically, storage location greatly implicates the development of SBL systems. The SP storage is limited to archive large data and would require external storage infrastructures. People will have separate opinions about facility of storing huge archives on a cloud-based storage. For some people, it is not acceptable because of private nature of personal lifelogs; whereas, others accept it if appropriate data hosting, backup, security, and retrieval facilities are ensured by a service provider. Practically, self-hosting of personal lifelogs is desirable by most lifeloggers with the choice of sharing a part of their digital memories securely.

However, sharing can have unclear consequences. A person might have control of choosing information to share but he/she would not be able to decide of how self-publicized information should be used and interpreted. Furthermore, user consent is also an important related challenge which needs great consideration while defining personal lifelog control, access, and monitoring policies [37]. The SBL systems should be developed by using privacy by design [38] framework, where privacy and data protection is considerably embedded in the entire development process and use of a system. However, privacy by design has received critics of being complex and affects functional requirements of a system [4]. Therefore, a tradeoff has to be settled between privacy and functionalities of SBL systems. More or less, privacy by design has to be incorporated in SBL systems development and to be tailored by the information retrieval developers while developing SP-based lifelog retrieval tools.

5.7.2 After-Logging

The personal lifelogs generated by SBL would remain within lifeloggers' ownership during their lifetimes. However, a legitimate question is why and how long a personal lifelog should retain when a data gatherer dies, commonly referred as "after-logging" challenges. The possibilities could be either the deletion of data or passing to a trusted relative. However, if passed then concerns arise about keeping a personal lifelog forever or for certain generations. A past person's personal lifelog could contain valuable historical contextual information and society may emphasize on keeping it for potential applications. The advanced storages can retain personal lifelogs indefinitely with low cost. However, a personal lifelog may contain a detailed trace of life, which may raise privacy concerns even when a lifelogger is dead.

5.7.3 Bystanders

The SP-based personal lifelog harnessing can potentially pose privacy challenges to the people around at the time of lifelogging, commonly referred as bystanders. The total capture lifelogging can create privacy consequences for the people beyond a lifelogger by capturing them without their consents. For example, using SP to record pictures, videos, and audios at a common place, can capture the presence of other people as well without their consents. This could produce severe consequences, if the captures containing some sensitive material are publicized (e.g., shared using a social media). Deleting the captures will not only limit relevant cues but also informative contents of personal lifelogs [39]. Thus, privacy challenges are not only from lifelogger perspectives but also from the bystanders encountered by a lifelogger.

5.7.4 Anonymisation

The SP is a highly portable device and has strong probability of being lost or damaged. A privacy concern arises when a SP with a rich personal lifelog is lost and it could be devastating for a lifelogger. In this and many other respects, the issue of anonymising personal lifelog in SP will receive research attentions. We believe that anonymisation should be implemented at access time, dynamic process, and dependent on user access policies instead of being non-reversible capture time process.

5.7.5 Ethics

Another key privacy issue is ethics of SBL. The information captured in a personal lifelog would not only be related to a lifelogger but would also contain abundance of information about people and situations happening around a lifelogger spatially and temporally. For example, if someone is captured while practicing a crime then the lifelogger is obliged to report it? It is very crucial to critically highlight and analyze the ethical, legal, and social issues, which may arise from SBL.

5.8 Tasks Standardization

The SBL is not standardized due to lacking of considerable systematic research efforts. To-date, a specific methodology of SBL for lifelog generation is not defined, which show essential components, sequencing and interactions of components, inputs and outputs of events, description of functions of components, identification of usable technologies, etc. The lack of standardization can be due to a number of reasons:

5.8.1 Data Collection Technologies Immaturity

A SP has a rich set of sensors to capture contents and contextual information [15]. However, they are limited to capture dynamic information from different aspects of a person's life. For example, to capture a person's healthcare information one cannot rely on SP sensors only but also has to use external specialized health sensors. The integration of every possible sensor in SPs is not possible due to size problem and their usage could have severed affects on SP operations such as depleting battery power, jeopardizing SP normal operations, etc. Furthermore, some of the SP sensors (e.g., camera) are far behind in performance than commercial sensors enabled devices. The wearable lifelogging has made improvements by building specialized lifelogging devices such as SenseCam, Memeto, OMG Autographer, etc., which have valuable lessons for SBL systems developers.

5.8.2 Human Life Dynamics

New requirements are taking birth due to human life dynamics. A rich prosthetic memory would be constructed if everything about a person's life is captured in real-time. Technically, SBL systems are using inertial sensors and body worn sensors to capturing a subset of life events and activities. However, they cannot scale dynamically for new events to constitute real reflection of a person's life experiences. Extreme SBL can solve this problem. However, the approach is not feasible and practical due to a number of reasons to-date: (1) requiring extreme use of additional technologies, which are not possible to be predicted that what would be integrated in the future SPs, (2) SBL applications predicting tomorrow's new activities and events are not technologically feasible to be built at present, and (3) peoples' interests in daily life events vary considerably where some events would be of high worth than others. However, semantically organizing and interlinking information in personal lifelogs can infer identification and handling of future events information using existing information. Similarly, using machine learning techniques can help in identifying patterns in personal lifelogs for predicting future events and activities.

5.8.3 Test Collection and Evaluation

The SBL research is lacking with availability of lifelog datasets and reference test beds. The available systems have developed their custom methodologies for test collections and results evaluations. However, none of the test collections is publicized for using in other researches' evaluations and ensuring integrity of the results. Gathering a dataset of real-time contexts and events information is a complex process and may encounter a number of problems. Firstly, developing a robust application to run for hours on SP requires valued programming skills and experiences. Secondly, technical (i.e., both hardware and software) issues may hinder problems in data collection. Thirdly, organizing and collecting an effective group of participants' sample from a population is a tedious and time consuming task. Finally, logging life experience requires dedication, devotion, and active participants' involvement and reporting. It should be ensured that participants may not report sparse or erroneous information due to not charging or forgetting to carry their SPs. Publicizing lifelog datasets would enable repeating of a lot of experiments for accurate determination of progress in SBL research. The people may be reluctant about publicizing them lifelog datasets due to privacy issues as they may have logged their personal contextual information, whose potential misuse could be devastating and may raise legitimate question about their security. However, anonymised versions of lifelog datasets could be helpful for SBL research for comparing the existing systems on the same test bed. The experiences and lessons from Reality Mining and CRAWDAD initiatives could be used in this respect.

5.8.4 Information Retrieval

The SBL systems are supposed to provide omnipresent access to lifelogs for retrieving required and related events information spatially and temporally. The SBL researchers have developed very basic types of information retrieval methodologies by displaying information in time line or clustering information into events and sub-events using time and location information. The searching and mining in such organizations is tedious and retrieving a particular piece of information would require one to review all of the events of a day, at least. More sophisticated types of information retrieval methods can be developed by leveraging and learning lessons from the information retrieval technologies developed for semantic memory retrieval domain and Human-computer Interaction (HCI). However, significant research is needed in SBL to clearly describe that how and what far traditional semantic memory retrieval techniques and others can be refined and made applicable for searching and retrieving SBL lifelogs.

6. CONCLUSIONS

The technological advancements have made smartphone as a de-facto lifelogging device. The sensory, processing, storing and networking capabilities enables smartphone to capture and store a wide variety of effective information about almost all aspects of our lives as well as contextual and environmental information. Realizing the fact, researchers have come up with SBL systems and have fostered SBL research as a mainstream research activity to solve real-world problems which would be impossible otherwise. Despite of all, SBL research is still infancy and faces with several issues and challenges regarding capturing, processing storage, ownership, privacy and security, and retrieval and visualization, which can impede its widespread adaptation of SBL. In this paper, we have thoroughly analyzed the SBL researches from different aspects and have listed a number of SBL research gaps, and issues and challenges to provide research opportunities for the researchers. We hope that identification of issues, challenges will provide an instrument, which will aid further into the development of this promising technology.

6. REFERENCES

1. Itzgibbon, A, & E. Reiter, Memories for life: Managing information over a human lifetime. *UK Computing Research Committee Grand Challenge proposal*, 22: 13-16 (2003).
2. Dodge, M, & R. Kitchin, Outlines of a world coming into existence: pervasive computing and the ethics of forgetting, *Environment and Planning B: Planning and Design* 34: 431-445 (2007).
3. Mohamed, E.S.T. Designing and evaluating a user interface for continues embedded lifelogging based on physical context. *Ph.D Thesis, School of Computing Science, Newcastle University, Newcastle*: (2012).
4. Gurrin, C., A.F. Smeaton, & A.R. Dohetry. LifeLogging: Personal Big Data. *Foundations and Trends in Information Retrieval*, 8: 1-125 (2014).
5. Kalnikaite, V., A. Sellen., S. Whittaker. & D. Kirk. Now let me see where i was: understanding how lifelogs mediate memory. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: 2045-2054 (2010).
6. Sellen, A, & S. Whittaker. Beyond Total Capture: A Constructive Critique of Lifelogging. *Communications of the ACM*, 53: 70-77 (2010).
7. Hodges, S., L. Williams., E. Berry., S. Izadi., J. Srinivasan., A. Butler., G. Smyth., N. Kapur. & K. Wood. SenseCam: a retrospective memory aid. *Proceedings of the 8th international conference on Ubiquitous Computing*: 177-193 (2006).
8. Siewiorek, D., A. Smailagic., J. Furukawa., A. Krause., N. Moraveji., K. Reiger., J. Shaffer. & F. L. Wong. SenSay: A Context-Aware Mobile Phone. Proceedings of the 7th IEEE International Symposium on Wearable Computers: 248 (2003).
9. Shah, M., B. Mears., C. Chakrabarti. & A. Spanias. Lifelogging: Archival and retrieval of continuously recorded audio using wearable devices. I*EEE International Conference on Emerging Signal Processing Applications (ESPA)*: 99-102 (2012).
10. Chennuru, S. K., P. Chen., J. Zhu. & J. Y. Zhang. Mobile Lifelogger - Recording, Indexing, and Understanding a Mobile User's Life. I*nternational Conference on Mobile Computing, Applications, and Services*: 263-281 (2010).
11. Belimpasakis, P., K. Roimela. & Y. Yu. Experience Explorer: A Life-Logging Platform Based on Mobile Context Collection. *3rd Int. Conference on Next Generation Mobile Applications, Services and Technologies*: 77-82 (2009).
12. Albatal, R., C. Gurrin., J. Zhou., Y. Yang. & D. Carthy. SenseSeer Mobile-Cloud-Based Lifelogging framework. *The IEEE International Symposium on Technology and Society (ISTAS)*: 27-29 (2013).
13. Rawassizadeh, R., M. Tomitsch., K. Wac. & A. M. Toja. UbiqLog: a generic mobile phone-based life-log framework. *Personal Ubiquitous Comput*.17: 621-637 (2013).
14. Ali, S., S. Khusro., A. A. Khan. & L. Hassan. A Survey of Mobile Phones Context-Awareness Using Sensing Computing Research. *Journal of Engineering and Applied Sciences* 33: 75-93 (2014).
15. Ali, S., S. Khusro., A. Rauf. & S. Mahfooz. Sensors and Mobile Phones: Evolution and State-of-the-Art. *Pakistan journal of science* 66: 386-400 (2014).
16. Myka, A. Nokia lifeblog -- towards a truly personal multimedia information system. *Workshop on Mobile Data Banken and Information Systemes*: 21-30 (2005).
17. Memon, M. A., S, Bhatti. & N. A. Mahoto. A Digital Diary: Remembering the Past Using the Present Context. *Mehran University Research Journal of Engineering & Technology* 35: 275-286 (2016).
18. Ali, S. Exploiting Sensor Data Semantics for Smartphone-Based Lifelogging: Towards the Development of Digital Prosthetic Memory on Smartphone. P*hD Thesis, Department of Computer Science, University of Peshawar*: (2018).
19. Khan, I., S. Ali. & S. Khusro. Smartphone Based Lifelogging: Investigation of Data Volume Generation Strength of Smartphone Sensors. In *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*.
20. Song H., Jiang D. (eds) Simulation Tools and Techniques, Springer, Berlin, Germany, p. 63-73 (2019).
21. Wong, K. Y. Cell phones as mobile computing devices. *IT professional* 12: 40-45 (2010).
22. Barton, J. J., S. Zhai. & S. B. Cousins. Mobile phones will become the primary personal computing devices. *Proc. of 7th IEEE Workshop on Mobile Computing Systems and Applications*: 3-9 (2006).
23. Gemmell, J., G. Bell. & R. Lueder. MyLifeBits: a personal database for everything," *Commun. ACM* 49: 88-95(2006).
24. Ali, S. & S. Khusro. Mobile Phone Sensing: A New

Application Paradigm," *Indian Journal of Science and Technology* 9: (1-41) (2016).Gurrin, C., Z. Qiu., M. Hughes., N. Caprani., A. R. Doherty.m S. E. Hodges. & A. F. Smeaton. The Smartphone As a Platform for Wearable Cameras in Health Research. *American Journal of Preventive Medicine* 44: 308–313 (2013).

25. Aizenbud-Reshef, N., E. Belinsky., M. Jacovi., D. Laufer. & V. Soroka. Pensieve: augmenting human memory. *Extended Abstracts on Human Factors in Computing Systems*: 3231-3236 (2008).
26. Froehlich, J., M. Y. Chen., S. Consolvo., B. Harrison. & J. A. Landay. MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones. P*roc. of the 5th international conference on Mobile systems, applications and services*: 57-70 (2007).
27. Packer, H. S., A. Smith. & P. Lewis. Memorybook: Generating Narratives from Lifelogs. *Proceedings of the 2nd Workshop on Narrative and Hypertext*: 7-12 (2012).
28. Sravanthi, K.and T. S. Reddy. Applications of Big data in Various Fields. *International Journal of Computer Science and Information Technologies*: 4629-4632 (2015).
29. Sangkeun, L., G. Gihyun., H. Inbeom. L. Sang-goo. LifeLogOn: A Practical Lifelog System for Building and Exploiting Lifelog Ontology. *IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC)*: 367-373 (2010).
30. Ahmed, M., H. H. Hoang., M. S. Karim., S. Khusro. M. Lanzenberger., K. Latif., E. Michlmayr. K. Mustofa. H. T. Nguyen. & A. Rauber. SemanticLIFE'-A Framework for Managing Information of A Human Lifetime. *Sixth International Conference on Information Integration and Web Based Applications & Services*: (2004).
31. Chen, C. P. & C.-Y. Zhang. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences* 275: 314-347 (2014).
32. Rawassizadeh, R. A holistic multi-purpose life logging framework. *Ph.D Thesis, uniwien,* (2012).
33. Doherty, A. R. & A. F. Smeaton. Automatically segmenting lifelog data into events. *Ninth International Workshop on Image Analysis for Multimedia Interactive Services*: 20-23 (2008).
34. Yen, An-Zi., H. Huang. & H. Chen. Personal Knowledge Base Construction from Text-based Lifelogs. *Proc. of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval:* 185-194 (2019).
35. Liggins II, M., D. Hall. & J. Llinas. *Handbook of multisensor data fusion: theory and practice*. CRC press, (2017).
36. Zafeiropoulos, A., N. Konstantinou., S. Arkoulis., D. E. Spanos. & N. Mitrou. A Semantic-Based Architecture for Sensor Data Fusion. *2nd Int. Conf. on Mobile Ubiquitous Computing, Systems, Services and Technologies*: 116-121 (2008).
37. Jacquemard, T., P. Novitzky., F. O' Brolchain. A. F. Smeaton. & B. Gordijn. Challenges and Opportunities of Lifelog Technologies: A Literature Review and Critical Analysis. *Science and Engineering Ethics* 20: 379-409 (2014).
38. Langheinrich, M. Privacy by design—principles of privacy-aware ubiquitous systems. *International conference on Ubiquitous Computing*: 273-291 (2001).
39. Elagroudy,P., M. Khamis., F. Mathis., D. Irmscher., A. Bulling. & A. Schmidt. Can Privacy-Aware Lifelogs Alter Our Memories?. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*: (2019).
40. Vemuri,S., S. Chris. & B. Walter. iRemember: a Personal, Long-term Memory Prosthesis. *3rd ACM Workshop on Continous Archival and Retrieval of Persoanl Experiences CARPE'* 06: 65-74 (2006).