Life-Tags: A Smartglasses-based System for Recording and Abstracting Life with Tag Clouds

ADRIAN AIORDĂCHIOAE, MintViz Lab, MANSiD, University Ştefan cel Mare of Suceava, Romania
RADU-DANIEL VATAVU, MintViz Lab, MANSiD, University Ştefan cel Mare of Suceava, Romania

We introduce "Life-Tags," a wearable, smartglasses-based system for *abstracting life* in the form of clouds of tags and concepts automatically extracted from snapshots of the visual reality recorded by wearable video cameras. Life-Tags summarizes users' life experiences using word clouds, highlighting the "executive summary" of what the visual experience felt like for the smartglasses user during some period of time, such as a specific day, week, month, or the last hour. In this paper, we focus on (i) design criteria and principles of operation for Life-Tags, such as its first-person, eye-level perspective for recording life, passive logging mode, and privacy-oriented operation, as well as on (ii) technical and engineering aspects for implementing Life-Tags, such as the block architecture diagram highlighting devices, software modules, third-party services, and dataflows. We also conduct a technical evaluation of Life-Tags and report results from a controlled experiment that generated 21,600 full HD snapshots from six indoor and outdoor scenarios, representative of everyday life activities, such as walking, eating, traveling, etc., with a total of 180 minutes of recorded life to abstract with tag clouds. Our experimental results and Life-Tags prototype inform design and engineering of future life abstracting systems based on smartglasses and wearable video cameras to ensure effective generation of rich clouds of concepts, reflective of the visual experience of the smartglasses user.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; **Interactive systems and tools**; **Ubiquitous and mobile devices**; • **Software and its engineering**;

Additional Key Words and Phrases: Smartglasses; Wearable video cameras; Wearable computing; Tag clouds; Word clouds; Lifelogging; Experiment; Automatic tag extraction.

ACM Reference Format:
Adrian Aiordăchioae and Radu-Daniel Vatavu. 2019. Life-Tags: A Smartglasses-based System for Recording and Abstracting Life with Tag Clouds. *Proc. ACM Hum.-Comput. Interact.* 3, EICS, Article 15 (June 2019), 22 pages. https://doi.org/10.1145/3331157

1 INTRODUCTION

Digital lifelogging is becoming more and more accessible due to the prevalence and variety of wearable sensing devices, such as activity trackers embedded in smartwatches and fitness bands [30, 31], high resolution action cameras designed for sports and adventure enthusiasts [23], commercial products dedicated to lifelogging [41,43,53], and tiny video cameras embedded in smartglasses for everyday eyewear [47,52], but also due to personal, portable storage devices and cloud storage services becoming less expensive and accessible each day. According to Gartner's November 2018 report [18] on wearable electronic devices, worldwide shipments of wearables will reach 225

Authors' addresses: Adrian Aiordăchioae, MintViz Lab, MANSiD, University Ştefan cel Mare of Suceava, 13 Universitatii, Suceava, 720229, Romania, adrian.aiordachioae2@student.usv.ro; Radu-Daniel Vatavu, MintViz Lab, MANSiD, University Ştefan cel Mare of Suceava, 13 Universitatii, Suceava, 720229, Romania, radu.vatavu@usm.ro.

2573-0142/2019/6-ART15 $15.00
https://doi.org/10.1145/3331157

Fig. 1. As wearable video cameras, such as those embedded in smartglasses or video cameras that can be directly clipped onto clothes, become more and more accessible, they open up a world of opportunities for continuously recording every aspect of the visual reality that we experience. In this work, we address design and engineering aspects of generating "life tags," which are summaries of daily life presented to users as tag clouds. In this figure, a user wears our Life-Tags system with a camera embedded in the eyewear that delivers snapshots to the smartphone for further processing.

millions of units in 2019, including smartwatches, head-mounted displays, smart clothing, ear-worn, wristbands, and sports watch devices. The increasing availability and popularity of smart wearables that can record and store snapshots and videos of their users' lives combined with the growing interest in recording oneself and posting video content and photographs on social networks [10,16,29,54] will furthermore impact the adoption of lifelogging as a way of life.

However, lifelogging is known to generate large amounts of data, *e.g.*, the Autographer wearable video camera [14], now a discontinued product, could generate 1.1 million images per day, 479.6 GB in one year, and 40.8 TB in a lifetime; see Gurrin *et al.* [22, p. 34] for statistics regarding the size of visual lifelog data. From this perspective, efficient presentation and retrieval techniques are particularly important to make visual lifelogs effectively useful. To this end, previous work has introduced lifelog summaries as timelines of snapshots and concepts automatically extracted from recorded videos and presented to lifeloggers using graphical user interfaces on desktop PCs [26,35,61]. Although useful for performing browsing and query tasks, such as identifying similar events [35] or searching the lifelog according to various criteria [61], presenting lifelog data effectively to users is nevertheless challenging, a perspective from which it resembles very much digital tools for managing photography. Regarding the latter, in Donald Norman's [44] words *"It is relatively easy to take digital photographs, easy to share them from the display on the camera itself ... The design challenge is to keep the virtues while removing the barriers: make it easier to store, send, share. Make it easier to find just the desired pictures years after they have been taken and put into storage. These are not easy problems, but until they are overcome, we will not reap the full benefits of photography."* (p. 51). Therefore, despite the recent availability of lifelogging gadgets and applications and low storage costs for the lifelog data, innovations are still needed in terms of effective presentation of the lifelog to users. We believe that such innovations can come from the Engineering Interactive Computing Systems community through careful, dedicated consideration of the design and implementation of input/output devices, interaction techniques, user interfaces and, ultimately, complete interactive systems for lifelogging and abstracting life. Moreover, collection of snapshots and videos using wearable video cameras can create opposition from passersby, who may react to the invasion of their privacy [11,34], an aspect that further motivates research towards new, privacy-oriented forms for storing, presenting, and sharing the visual lifelog.

In this work, we propose *tag clouds* as an alternative representation for visual lifelog data, and we specifically address designing and engineering aspects of wearable, smartglasses-based systems that support passive, automatic capture of visual lifelogs with the purpose to *abstract life*. To this end, we focus on the implementation details of Life-Tags, our wearable system for the automatic acquisition and processing of snapshots of the visual life. Our practical contributions are as follows:

(1) We introduce "Life-Tags," a new concept for presenting lifelog data to users and a smartglasses-based implementation of a system that generates clouds of tags from visual concepts automatically extracted from snapshots captured by the video camera embedded in the smartglasses; see Figure 1 for an illustration of our wearable system in operation. Life-Tags is different from other lifelogging systems in at least two ways: (i) by abstracting life using clouds of tags and visual concepts instead of simply storing snapshots and videos, and (ii) due to its privacy-oriented operation as a direct consequence of employing tag clouds.

(2) We describe the technical aspects of the Life-Tags system: its design requirements and principles of operation, such as first-person, eye-level perspective and privacy-oriented operation mode and discuss the engineering and technical details of its implementation, such as the block architecture diagram of devices, software modules and components, third-party services, and dataflows that constitute the Life-Tags system.

(3) We also contribute a technical evaluation of Life-Tags in the form of a controlled experiment conducted to understand the effect of snapshots' sampling frequency on the richness of the visual lifelog and storage requirements. These results can be used to inform design and engineering of *any* wearable video camera-based system and, therefore, can be readily used by practitioners in their prototypes and systems for lifelogging and *abstracting life*. For instance, our results highlight the fact that current commercial products, such as the Narrative Clip 2 wearable video camera [43] (a camera so popular and in such high demand that at the time of April 2019 needed to temporarily break sales to catch up with production and shipment to be able to deliver orders to waiting customers[1]), fail to capture about 68% of the concepts describing the visual experience of the lifelogger, simply because they default to snapshotting life just twice per minute compared to sampling twice per second.

Our practical contributions, in terms of implementation details of our interactive system, from design requirements and principles of operation to technical engineering, measurements of performance, and reporting of experimental results, will benefit researchers and practitioners of the Engineering Interactive Computing Systems community by informing design and engineering aspects for future prototypes, applications, systems, and contexts of use for lifelogging.

2 RELATED WORK

We discuss in this section prior work on lifelogging and smartglasses-based systems and applications, and conclude with a summary of our practical contributions and how they relate to previous work.

2.1 LifeLogging

LifeLogging has been described as a phenomenon, whereby people digitally record their daily lives [22], as personal big data [22], but also as a form of pervasive computing [17], representing *"a unified digital record of the totality of an individual's experiences, captured multimodally through digital sensors and stored permanently as a personal multimedia archive"*; see Dodge and Kitchin [17, p. 431]. Any lifelogging system must include several components: a *sensor* that captures some aspect of the everyday life of the lifelogger, such as a wearable video camera [57,59,60] or an activity or location tracker [42,62]; *middleware software* that cleans the raw data and eventually

[1]Temporarily Out of Stock. Narrative Clip 2 is in high demand, http://getnarrative.com/shop

aligns multiple data streams delivered by different sensors; a *segmentation module* that structures the cleaned, aligned data into units called "events;" a *semantic analyzer* and *annotation component* that constructs the lifelog; and, finally, the *retrieval technique* and corresponding *user interface* for presenting the lifelog to the user. Gurrin *et al.* [22] discuss each component in great detail in their overview of lifelogging research and applications, providing many examples of how lifelogging systems have implemented these components in various ways. Notable examples of lifelogging prototypes from the scientific literature include SenseCam [26], the EyeTap technology [39], the DejaView system [15], and InSense [9], to name just a few. Also, several commercial products are available for lifelogging enthusiasts, such as the Narrative Clip 2 [43], MeCam [41], SnapCam [53], and Google Clips [13], among others.

Lifelogging occurs for a variety of desiderata, made possible by the potential of lifelog data to inform about how to live one's life [22]. Prior work on applications of lifelogging addressed "food-logging" [33], computer usage [58], sleep patterns [42], "wordometer" systems that compute an estimate of the number of words read in everyday life [5], monitoring aspects regarding the quality of life [62], vehicular lifelogging [2,40], or "thing-logging" for the Internet-of-Things [19]. For example, Kitamura *et al.* [33] implemented a system that could recognize images containing food with 88% accuracy, estimate the food balance with 73% accuracy, and present users with a visualization of their food log; and Zini *et al.* [62] proposed a system designed to monitor four distinct aspects of life quality, such as activities, sleep quality, level of fatigue, and mood, which were presented to users as gauge charts on their smartphones. An important application of lifelogging is to provide memory support or a "memory prosthesis," for example to people with Alzheimer's, enabling them to relive recent experiences by resorting to the lifelog [8,55]. Another example is Al Mahmud *et al.* [3], who designed an image capturing device for people with aphasia that took photographs and added tags automatically.

The typical day of the lifelogger is segmented by the lifelogging application into events that are usually presented along a timeline highlighting representative snapshots and metadata from the lifelog [22]. For example, the SenseCam viewer allows playback of snapshots on a PC and shows the corresponding sensor readings [26, p. 185]; and Lee *et al.* [35] presented the user with their day segmented into events represented by keyframes, while also retrieving and displaying similar events from the lifelog; see [35, p. 343] for a snapshot of the user interface. Zhou et al. [61] introduced LIFER, a search engine for lifelogging applications that implemented various search criteria, such as date and time, location, or activity to query the lifelog. After filling in the search criteria, matching data were presented to the user in the form of a list of snapshots with metadata.

2.2 Smartglasses-based Systems and Applications

There is a wide literature on smartglasses with topics ranging from technical prototyping [1,7,28] to applications [57,59,60], and to studies and investigations to understand privacy and ethical concerns regarding public video recording with personal, wearable video cameras [11,27,34]. For example, Yang *et al.* [57] developed a prototype for the visual augmentation of daily life, providing users with a refined visual recognition result computed from multiple input images; Zhao *et al.* [60] introduced CueSee, a system that enabled people with low vision to locate a specific product on a grocery store shelf independently and efficiently; and the ForeSee system of Zhao *et al.* [59] was designed to improve the visual experience of users with low vision by providing magnification, contrast and edge enhancement of the visual reality. Other applications of smartglasses include vision enhancement during night conditions [28], providing assistance to colorblind people, such as the Colorizer system [46], or helping with obstacle detection [1,7]. Smartglasses that embed tiny video cameras are specifically appealing for implementing lifelogging systems and applications as they capture snapshots and videos from the eye level perspective.

2.3 The Contributions of Life-Tags to the Scientific Field and Community of Engineering Interactive Computing Systems

In this work, we introduce a new concept to present visual lifelog data to users in the form of word clouds of tags automatically extracted from snapshots captured by the video camera embedded in smartglasses. To this end, we present the Life-Tags interactive prototype and discuss its design requirements, principles of operation, implementation aspects, and interaction model. Our approach focuses on the new concept of *abstracting life* by providing users with executive summaries of what their visual experiences were like, as recorded from the first-person, eye-level perspective enabled by video camera-based smartglasses, rather than with a list of snapshots and videos.

We position our work in the literature of lifelogging research, although our approach is directed towards *abstracting* rather than *logging*. From this perspective, Life-Tags enables lifeloggers with a new means to store and share their life experiences that is privacy-oriented and, therefore, likely more acceptable for displaying on social networks, online personal blogs, or public screens designed to render new forms of ambient media [45,56]. We also position our work in the broad literature of engineering interactive computing systems and, especially, wearable computers and interactive devices, as we are interested in practical engineering aspects for implementing life abstracting systems and applications, which can serve to inform future designs of interactive wearable lifelogging systems, *e.g.*, How many snapshots are needed to generate a comprehensive word cloud of concepts to describe the visual reality? or How to balance the richness of the tag cloud and the memory needed to store the snapshots from which tags were extracted?

In particular, we focus on a timely research question for lifelogging systems that has not been explored for abstracting life: *how much to log and sample from the visual world*. For example, in their survey of lifelogging as personal big data, Gurrin *et al.* [22] provided many examples of lifelogging methods, devices, and systems, such as the example of Richard Buckminster Fuller, who manually logged every 15 minutes of his activity between 1920 to 1983 into a scrapbook [22, p. 5]. Another example is JenniCam [24] that live streamed college student Jennifer Kaye Ringley's daily activities on the web at a frequency rate of one snapshot every 3 minutes. Today's high-speed action cameras can capture 1080p quality images at impressive speeds of 120 fps or 240 fps; see Hall [23] for an overview of the ten action cameras for 2019. Of course, these examples represent extreme cases of sampling life, but the frequency of sampling does have an impact on the size of the visual lifelog as well as on the richness of the data presented to the lifelogger and, consequently, it is important for practitioners to understand the effects of this variable when creating their prototypes and interactive systems. In between these extremes lie commercial lifelogging cameras, such as the Narrative Clip 2 [43], a 5 Mp camera that takes one snapshot every thirty seconds, or Google Clips [13], a video camera that automatically captures short clips of several seconds when it detects people or pets. Unfortunately, the question of how much to sample the visual world has not been addressed for abstracting life, despite it being a valid and important engineering aspect for lifelogging systems based on wearable video cameras. Next to our technical description of the design and engineering aspects of the Life-Tags system, we also investigate in this work the practical effect of snapshots' sampling frequency on the lifelog data collected by video camera-based systems with the goal to inform and guide design of future interactive prototypes, devices, applications, and systems to abstract visual life effectively.

3 LIFE-TAGS

We describe in this section the implementation of Life-Tags, our smartglasses-based system designed to automatically capture snapshots using the video camera embedded in the smartglasses and to abstract life in the form of word clouds of visual concepts; see Figure 2 for examples of tag clouds generated by Life-Tags. The snapshots are stored temporarily on the user's smartphone, offloaded to a permanent storage, such as a desktop PC or the cloud, and fed into a third party processing service to generate descriptive tags, which are used to compile word clouds.

Fig. 2. Tag clouds generated by Life-Tags for six scenarios. From left to right and top to bottom: *bar*, *building*, *bus*, *train*, *supermarket*, and *restaurant*. Note the power of abstracting life using simple concepts, *e.g.*, the bus traveling experience (top right cloud) can be described using concepts such as "snow", "vehicle", "road", "winter", etc., reminding the user of a cold winter afternoon.

3.1 Design Criteria and Principles of Operation for Life-Tags

We adopted the following six design criteria (D_1 to D_6) for Life-Tags by surveying the literature on lifelogging, focusing on the unique features of smartglasses with embedded video cameras, and by identifying requirements for systems designed to abstract life:

D_1. *Passive and automatic lifelogging.* Life-Tags should automatically capture and tag snapshots without the user's explicit or conscious initiative. This principle of "always on" operation relieves the user from the need to actively monitor and control what is being captured and when and, consequently, Life-Tags does not interfere with the life events experienced by the user. From this perspective, Life-Tags falls into the category of *"passive visual capture"* lifelogging devices and systems according to Machajdik *et al.* [37].

D_2. *First-person perspective and eye-level point of view.* Life-Tags should capture the user's point of view when recording snapshots of the visual reality and, therefore, the video camera should be positioned at eye level. Consequently, the lifelog will present users with the same perspective of the visual world as experienced in the first place. This design criterion and principle of operation, although common sense, seems to be ignored by many commercial wearable cameras that were designed to be worn by clipping onto clothes [13,41,43,53]. Nevertheless, the importance of a correct perspective and point of view has been explicitly stated in the scientific literature; see Gurrin *et al.* [22, pp. 20, 31, and 39] for a discussion.

D_3. *Inconspicuousness, yet privacy-oriented.* Life-Tags should not draw unwanted attention to the lifelogger. We target inconspicuousness at two levels: the wearer and the passersby. For the wearer, inconspicuousness prevents Life-Tags to interfere with everyday life activities and translates to always-on, passive operation. For passersby, it prevents unwanted opposition or manifestation of concerns [11], which should nevertheless be managed appropriately by the wearers of Life-Tags. Previous studies showed that lifeloggers are concerned about and take measures regarding the privacy of bystanders [27]. As a minimum requirement, Life-Tags should enable users to easily switch recording on and off; see Figure 6, left for a screenshot of Life-Tags' mobile user interface for the smartphone featuring a prominent Start/Stop button and clear indication of recording and operation status. This principle of operation connects to the recommendation *"communicate the intention of use"* of Koelle *et al.* [34] as well as to Chowdhury *et al.*'s [11] suggestion *"if a bystander inquires about the device, first the image capture must be paused, and then the objectives of our study must be explained"* (p. 15:2).

D_4. *Supporting memory access at various levels of granularity.* Life-Tags should support memory access for various user needs, from an abstract of visual concepts in the form of word clouds to access to actual snapshots [26,35,61] and video montages pertaining to the same activity, event, or concept. Figure 6, right presents a screenshot of Life-Tags' desktop user interface that enables visualization of individual snapshots and video montages of those snapshots as well as generation of word clouds. This design criterion was inspired from Sellen and Whittaker's [50] "five R's of memory access," which characterize distinct access requirements for lifelogs, *i.e.*, *recollecting*, *reminiscing*, *retrieving* information, *reflecting*, and *remembering* intentions (p. 76) to outline the potential benefits of lifelogging as memory support. For example, Life-Tags should support *recollecting* by enabling users to mentally relive life experiences triggered by familiar concepts, *e.g.*, "driving", "university", "walk", etc., of which some concepts could even lead to emotional responses (a transition from *recollecting* to *reminiscing*) when triggered by key tags, such as "graduation", "family", "ceremony", etc. Also, since Life-Tags operates with tags directly, a concept search facility can implement *retrieving* based on keywords; see Figure 6, right for the Life-Tags desktop user interface. The frequency of tags can activate a process of *reflection* and examination of life patterns, *e.g.*, the tags "indoors", "no person", "room", and "architecture" of the *building* scenario (see Figure 2, top middle) are powerful triggers to suggest the need for taking some time off and focusing on personal life. Finally, tag clouds can help *remembering intentions*, which are future actions to which we committed to and of which we are reminded by similar actions from our past captured in the tag cloud, *e.g.*, the words "shopping" and "market" (Figure 2, bottom middle) are reminders of periodic chores.

D_5. *Easy integration with other devices.* Although wearable video cameras feature reasonable memory storage (*e.g.*, the Narrative Clip has 8 GB memory, which can be used to store 4,000 photos or 80 minutes of video [43]), offloading storage to smart devices, such as smartphones, desktop PCs, or to the cloud is recommendable practice. According to this requirement, Life-Tags should easily integrate with a smartphone using a wireless connection.

D_6. *Easy integration with personal web logs and social networks.* Life-Tags should provide an effective, anonymous summary of the lifelog data to share on personal web logs or social networks, an ample social phenomenon that cannot be ignored [25,49], such as in the form of a word cloud describing the day of the lifelogger.

3.2 Implementation

Figure 3 shows a diagram block of the Life-Tags system highlighting devices, components, third-party services, and dataflows. We implemented Life-Tags using the SS-IP13 eyeglasses[2] that feature a full HD micro video camera (criteria D_2 and D_3), Wi-Fi operation, and a 90° field of view. The eyeglasses connect to a smartphone Android application (D_5) using the built-in Wi-Fi connection, and snapshots are captured at a resolution of 1920×1080 pixels via HTTP requests and are stored as JPEG files on the smartphone with a compression ratio of about 23:1. Figure 4 presents a code

[2]SS-IP13 is a local vendor name for the NorthVision Technologies glasses spy camera, http://northvisiontec.com/products/camera-spy/glasses-eyewear-camera/nc-c05glasses-camera19201080-avi-tf-card-videophoto-876.html

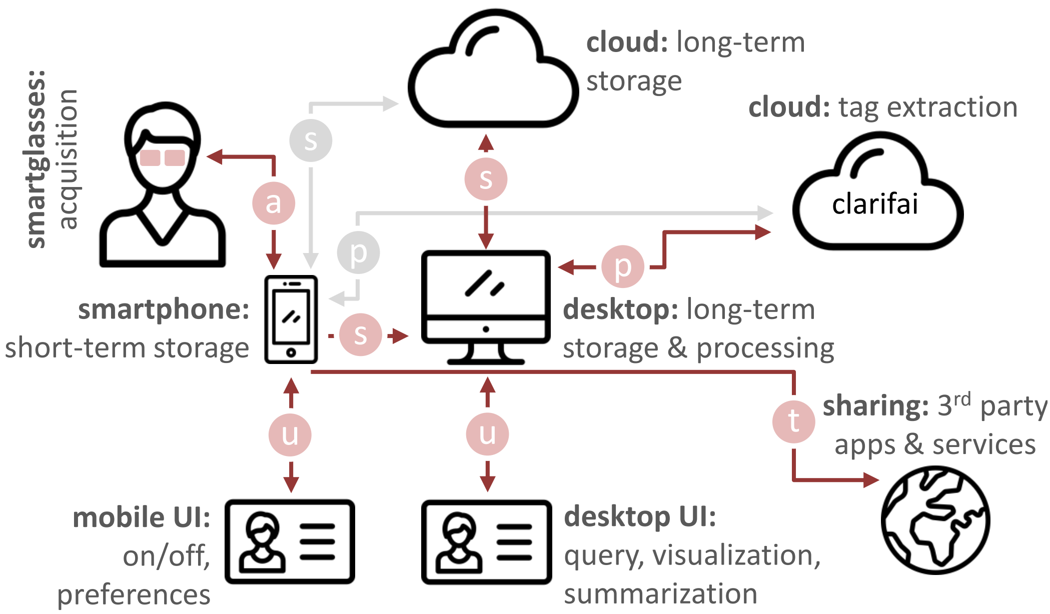

Fig. 3. Block architecture diagram of the Life-Tags system showing devices, components, third-party services, and dataflows. *Notes:* arrows indicate five dataflow types: acquisition (a), storage (s), processing (p), user interface (u), and sharing via third-party services (t); dark arrows illustrate default operation modes for Life-Tags, while light arrows show alternative dataflows.

snippet (Java for Android) illustrating the implementation of the snapshot acquisition process (dataflow (a) from Figure 3) from the video camera eyeglasses via a request-response procedure over `HTTP` running on a background, non-UI thread. The Android application features start and stop functionality (D_3) and, when on, performs automatic collection and storage of snapshots (D_1) at a frequency of two snapshots per second (2 fps), a default setting for Life-Tags at which we arrived through a trial and error process[3]; see Figure 6a for a screenshot of the mobile user interface for the smartphone app. Periodically, snapshots are offloaded to a desktop PC and, optionally, to a cloud storage (dataflows (s) in Figure 3), from where they are sent to `Clarifai` [12], a third-party cloud-based service for automatic description of images. Figure 5 presents a short code snippet (C#, .NET Framework) that implements dataflow (p) from Figure 3. Note that several cloud-based services are available for the automatic extraction of concepts from images, such as Google Vision AI [21], Amazon Rekognition [4], or AYLIEN [6]. However, since our goal with Life-Tags was to demonstrate the concept of abstracting life with a functional prototype rather than to compare existing systems or to innovate in terms of automatic tag extraction from images, we simply chose one of these third-party services to use in our implementation. Optionally, snapshots can be sent to `Clarifai` directly from the smartphone, which enables a tag cloud to be created directly on the mobile device. The service responds with a `JSON`-formatted text containing a list of identified concepts and their confidence values; see Figure 7 for an example showing concepts with confidence greater than 95%. The `JSON` message is parsed and a life abstract is generated (D_5) with various levels of interactivity (D_4), from a simple image of a word cloud to be shared online (D_6) to links to

[3]Capturing snapshots at higher rates on the smartphone resulted in missing data because of the multiple processing steps involved, *i.e.*, perform request over Wi-Fi, receive the incoming stream of bytes, and store the image on the SD card.

snapshots and video montages pertaining to specific concepts. Figure 6 illustrates screenshots of the graphical user interface of the mobile and desktop applications for controlling the operation of Life-Tags, sharing lifelog data, performing queries, visualizing snapshots, creating word clouds of tags and visual concepts from snapshots, and compiling summary videos.

```java
// Snapshots are captured at 2fps
new Timer().scheduleAtFixedRate(new TimerTask() {
  @Override
  public void run() { new GetSnapshot().execute(); }
  }, 0, 500 // 500ms approx. 2fps
);

// Acquire snapshots via HTTP
class GetSnapshot extends AsyncTask<Void, Void, Void> {
  ...
  // request a snapshot from the eyeglasses on a background, non-UI thread
  @Override
  protected Void doInBackground(Void... params) {
    HttpURLConnection urlConnection = null;
    try {
      URL url = new URL("http://192.168.10.1/snapshot.cgi?user=admin&pwd=");
      urlConnection = (HttpURLConnection)url.openConnection();
      urlConnection.connect();
      snapshotBytes = ByteStreams.toByteArray(urlConnection.getInputStream());
      if (isRecording)
        saveImage(Long.toString(System.currentTimeMillis()), snapshotBytes);
    }
    catch (MalformedURLException e) { e.printStackTrace(); }
    catch (IOException e) { e.printStackTrace(); }
    finally { if (urlConnection != null) urlConnection.disconnect(); }
    return null;
  }
}
```

Fig. 4. Code snippet (Java for Android) implementing snapshot acquisition from the video camera eyeglasses via `HTTP` requests performed on a background thread (dataflow (a) from Figure 3).

```csharp
private async Task ExtractConceptsFromSnapshot(byte[] snapshotBytes) {
  try {
    ClarifaiResponse<ClarifaiOutput<Concept>> response =
      await (new ClarifaiClient(API_KEY)).PublicModels.GeneralModel
      .Predict(new ClarifaiFileImage(snapshotBytes)).ExecuteAsync();

    if (response.IsSuccessful)  {
      // process concepts with confidence of at least 95%
      List<Concept> concepts95 = response.Get().Data.Where(c=>c.Value >= 0.95m).ToList();
      foreach (Concept concept in concepts95)
        ProcessConcept(concept.Name, concept.Value);
    }
  }
  catch (Exception exc) { /* exception handling */ }
}
```

Fig. 5. Code snippet (C#, .NET Framework) implementing the communication with the `Clarifai` service via an `HTTP` asynchronous request-response process (dataflow (p) from Figure 3).

The three core elements listed by Gurrin *et al.* [22] for lifelogging systems (Core$_1$ to Core$_3$) are implemented by Life-Tags as follows:

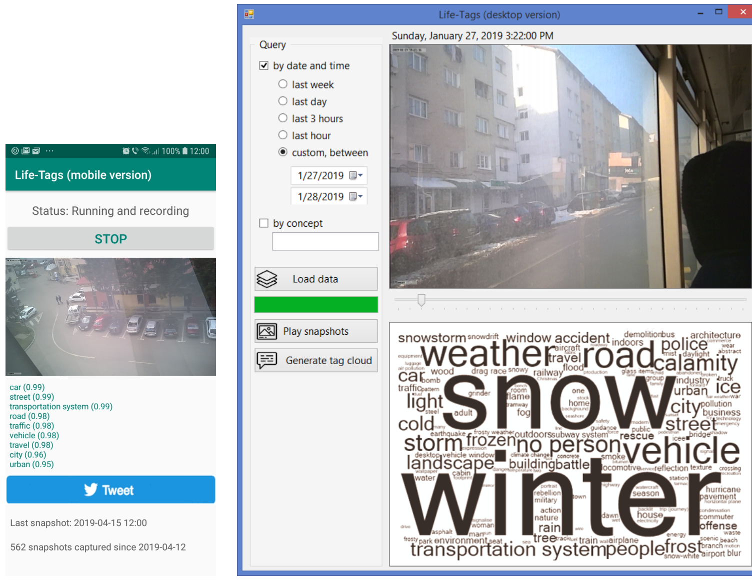

Fig. 6. Screenshots of the Life-Tags user interface for the smartphone (left) featuring start/stop functionality and status information; and for the desktop (right): running queries on the lifelog data, visualizing snapshots and videos, and creating word clouds of tags and concepts.


```
{
  "concepts":[
    {"id":"ai_GjVpxXrs", "name":"street", "value":0.98922026, "app_id":"main"},
    {"id":"ai_WBQfVV0p", "name":"city", "value":0.9872322, "app_id":"main"},
    {"id":"ai_w0Drrqn6", "name":"accident", "value":0.98443806, "app_id":"main"},
    {"id":"ai_TZ3C79C6", "name":"road", "value":0.9841032, "app_id":"main"},
    {"id":"ai_6lhccv44", "name":"business", "value":0.9829799, "app_id":"main"},
    {"id":"ai_rsX6XWc2", "name":"building", "value":0.98082936, "app_id":"main"},
    {"id":"ai_CpFBRWzD", "name":"urban", "value":0.9771112, "app_id":"main"}
  ]
}
```


Fig. 7. Example of (an excerpt from) a JSON message containing a list of concepts received from the tag extraction service for a snapshot from the *bus* scenario. In this example, seven concepts are listed in decreasing order of their confidence values, from 98.9% for "street" to 97.7% for "urban."

CORE$_1$. *Lifelogging*, *i.e.*, the process of passively gathering, processing, and reflecting on life experience data, is implemented in Life-Tags by collecting snapshots of the visual reality using the micro video camera embedded in the SS-IP13 eyeglasses.

CORE$_2$. *The lifelog*, *i.e.*, the data collected on the hard drive, in the cloud, or on a portable storage device, is represented by the individual snapshots stored in binary form. From these snapshots, word clouds are generated using tags automatically extracted by the external cloud service; see Figure 2 for a few examples of tag clouds.

CORE$_3$. *The surrogate memory*, *i.e.*, the data from the lifelog and the associated software to organize and manage lifelog data, is represented by the tag clouds generated on demand to abstract the visual life experienced by the lifelogger during some given time period. Life-Tags enables generation of tag clouds for (i) *custom time intervals*, *e.g.*, from January 1st to January 31st, 2019, and for (ii) *recent experiences*, *e.g.*, the last hour, last day, etc.

4 EXPERIMENT

We conducted a technical evaluation of Life-Tags in the form of a controlled experiment to understand and inform implementation aspects regarding the sampling frequency recommendable for the automatic capture of snapshots when abstracting life with tags. To this end, we considered two methods for sampling snapshots: (i) *uniform sampling* at a constant rate, such as one snapshot captured every 30 seconds [43], and (ii) sampling based on *event detection*, such as motion detected between two consecutively captured snapshots [26].

4.1 Design

Our experiment was a repeated-measures design with one human subject only (see the next subsection for a description of the participant) and the following three independent variables:

IV_1. Scenario, nominal variable, represents an everyday life scenario for which the abstract of the life experience is desired. We selected six distinct conditions for the Scenario variable: (1) commuting by *bus*, (2) walking inside a *building*, (3) having lunch at a self-service *restaurant*, (4) spending time in a *bar*, (5) traveling by *train*, and (6) grocery shopping at a *supermarket*. Each scenario was recorded for 30 minutes at the default sampling rate of Life-Tags of 2 fps, generating a total number of 3,600 full-HD images per scenario and, overall, 6 (scenarios) × 3,600 = 21,600 snapshots representing 180 minutes of recording.

IV_2. Sampling-Frequency, ratio variable, represents the sampling rate at which snapshots are captured by the wearable video camera. We controlled Sampling-Frequency using eight conditions: one snapshot captured every 0.5 seconds, 1 second, 2, 5, 10, 15, 30, and 60 seconds. These conditions include default settings for commercial wearable cameras, such as Narrative Clip 2 [43], which captures one snapshot every 30 seconds.

IV_3. Motion-Threshold, ratio variable, represents the threshold according to which a motion detection event is reported by comparing two consecutively captured snapshots. We considered five conditions for the Motion-Threshold variable, *i.e.*, 10%, 20%, 30%, 40%, and 50% of the pixels from the two frames need to change significantly in order for a motion event to be reported. We included this independent variable in our experiment design because some lifelogging cameras take snapshots based on various events, such as SenseCam [26] or Google Clips [13], instead of uniform sampling implemented by the Narrative Clip [43]. Motion detection was implemented using the `MotionDetector` class of the `Accord.NET` framework,[4] and was performed after the experiment using the complete dataset of all the snapshots that were collected at the default recording speed of Life-Tags of 2 fps.

4.2 Participants

We asked one volunteer, referred to in this paper with the fictive name Andrew, to wear the Life-Tags system and use it to lifelog each condition of the Scenario variable. Andrew, a 29-year old male, has a technical training in Computer Science and his life habits include working at a computer as part of his job (for approximatively 4-6 hours per day), walking, watching television, and exercising. He uses a smartphone and laptop on a daily basis. At the moment of the study, Andrew worn prescription glasses (+1.25 diopters for the left eye and +0.25 diopters for the right eye) and had the dry eye syndrome, for which he was taking medical treatment.

It is important to note that one human subject is enough for us to collect the necessary amount of data to perform the technical evaluation of Life-Tags in terms of understanding the effective sampling frequency to abstract life. The need for the human subject is simply to wear the system

[4] http://accord-framework.net/docs/html/T_Accord_Vision_Motion_MotionDetector.htm

in various experimental conditions. Actually, for the purposes of this technical evaluation, the conditions of the SCENARIO variable represent the actual "subjects," while the snapshots collected in each scenario represent "repeated trials," according to standard terminology employed for the design of controlled experiments. From this perspective, our experiment can also be seen as a within-subject design, as measurements are taken repeatedly over subjects (which are the scenarios) that go through several experimental conditions regarding sampling frequency and motion threshold. Moreover, studies with one participant only are not uncommon for evaluating lifelogging systems. For example, Zhou *et al.* [61] employed one lifelogger to test their LIFER search engine for lifelog data; Lucero [36] presented an autoethnography of one person's experiences living without a mobile phone; Shinohara and Tenenberg [51] conducted an observational study of one blind college student interacting with various technologies within her home; and Microsoft's SenseCam lifelogging device [26] was originally evaluated with just one patient with amnesia.

4.3 Measures

We evaluate the technical performance of Life-Tags using the following three measures representing our dependent variables (DV_1 to DV_3):

DV_1. NUM-CONCEPTS represents the number of distinct concepts or tags that are identified from the snapshots collected in a given scenario. The number of concepts depends on (a) the accuracy of the computer vision algorithm that processes the snapshots, (b) the duration of the recording (*i.e.*, longer recordings result in more snapshots and likely in more distinct concepts), but also (c) on the confidence level employed for the concept extraction procedure (*i.e.*, a lower level of confidence, such as 80%, will lead to more concepts being reported compared to a more conservative level, such as 95% or 99%). Therefore, the number of concepts can be formalized as a mathematical function of three variables:

$$\textsc{Num-Concepts}(\alpha, N, c) : \mathcal{A} \times \mathbb{N} \times (0..100] \rightarrow \mathbb{N} \tag{1}$$

where α is the computer vision algorithm used to extract tags and concepts given the set $\mathcal{A}$ of all possible algorithms, N is a positive integer representing the number of collected snapshots, and c is the confidence level, between 0% and 100%. In this work, we employ the `Clarifai` [12] online service to implement α with a confidence level of at least 95%, representing high confidence in the extracted concepts, and we compute NUM-CONCEPTS for all the 30 minutes recorded in each scenario. For brevity, we use the short notation NUM-CONCEPTS to actually denote NUM-CONCEPTS(`Clarifai`, 3600, 95%). For example, from the 3,600 snapshots collected in the *bus* scenario, a number of 30,475 tags were extracted with 95% confidence representing an average of approximately 8 high-confidence tags per snapshot. From these, 176 tags were distinct. Therefore, NUM-CONCEPTS is 176 for the *bus* scenario. The next section presents experimental results for this measure but, overall, a larger number of high-confidence concepts means better performance for the life abstracting system as a result of a richer capacity to capture, index, and describe the wearer's visual experience.

DV_2. LIFELOG-SIZE represents the total amount of bytes that are needed to store the snapshots collected in each scenario. LIFELOG-SIZE depends on the resolution at which snapshots are recorded and on the image compression algorithm used to store the snapshots. From this perspective, LIFELOG-SIZE is a function of two variables, *i.e.*, LIFELOG-SIZE(res, β), where res is the image resolution in megapixels and β the compressor. In this work, we use full-HD snapshots ($1920 \times 1080 = 2.1$ Mp), which is the maximum resolution supported by the SS-IP13 eyeglasses device, and we store snapshots compressed as `JPEG` files. For brevity, we use the short name LIFELOG-SIZE to denote LIFELOG-SIZE(2.1, `JPEG`). For example, the size of

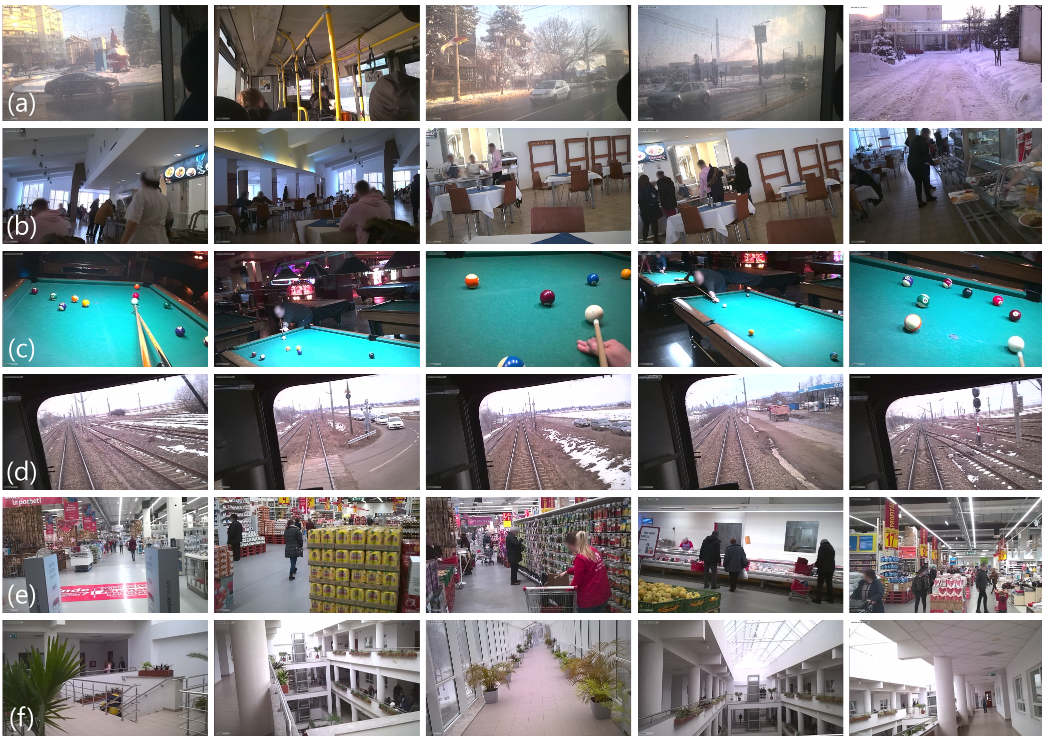

Fig. 8. Examples of snapshots captured in each scenario. From top to bottom: (a) commuting by *bus*, (b) having lunch at a self-service *restaurant*, (c) playing a pool game in a local *bar*, (d) traveling by *train*, (e) grocery shopping at a *supermarket*, and (f) walking inside a *building*.

the snapshots collected in the *bus* scenario varied between 150.9 KB and 645.7 KB, with an average size of 337.2 KB computed across all the 3,600 *bus* snapshots. A smaller size of the lifelog data is desirable given the same quality of the stored images.

DV_3. Tag-Cloud-Efficiency, defined as the ratio between the number of distinct concepts extracted from snapshots and the memory needed to store those snapshots:

$$\textsc{Tag-Cloud-Efficiency}(\alpha, N, c, res, \beta) = \frac{\textsc{Num-Concepts}(\alpha, N, c)}{\textsc{LifeLog-Size}(res, \beta)} \tag{2}$$

In this work, we compute Tag-Cloud-Efficiency for all the 3,600 snapshots of each scenario, extract tags with a confidence of 95%, and use full HD resolution and JPEG compression. Thus, the short name Tag-Cloud-Efficiency actually denotes Efficiency(`Clarifai`, 3600, 95%, 2.1, `JPEG`). For example, the abstracting efficiency in the *bus* scenario is 176/(3600 · 337.2 KB / 1024) = 0.148 distinct concepts per 1 MB of storage when sampling two snapshots every second. Efficiency combines the previous two measures and delivers a single performance criterion: a large number of high-confidence concepts that are extracted from a lifelog dataset consisting of snapshots with little memory requirements means better performance.

4.4 Task, Apparatus, and Scenarios

Andrew was asked to wear the Life-Tags system, which featured the SS-IP13 smartglasses with a micro video camera embedded in the frame, almost impossible to spot by the uninformed passersby; see Figure 1 from the Introduction. The Life-Tags application for Android was installed on his smartphone. Andrew was instructed to use the application to record each scenario for a total duration of 30 minutes. All scenarios considered, 180 minutes were recorded at the default sampling rate of 2 fps. A brief description of each scenario, as provided by Andrew, is given in the following:

(1) *Commuting to work using the bus.* In this scenario, Andrew lifelogged the trip from his home to the workplace, a journey during which he traveled 7 km by bus for approximately 23 minutes, and 700 m on foot, representing a 7-minute walk. Lifelogging started at approximately 15:00, when it was still daytime (sunset occurred at about 17:00). The season was winter and it had snowed heavily a few days before the recording took place. While for the first part of his journey, Andrew was sited and recorded the action mainly taking place outside the bus window, he was actively walking during the second part, having access to a larger field of view around him. The temperature during that afternoon was a few degrees below zero Celsius. Figure 8a illustrates a few snapshots for this scenario.

(2) *Having lunch at a restaurant.* Andrew had lunch at a self-service restaurant during 12:50 and 13:20. Activities included browsing the food-serving counter, engaging with the personnel, searching for a free table to sit down, eating, watching other people having lunch, and finally leaving the restaurant; see Figure 8b. The location was indoor and was reasonably well-lit.

(3) *Local bar.* In this scenario, Andrew went to a bar and played a game of pool with a friend. The location was indoor, poorly lit from just the dim lights installed above the pool tables. The bar was almost empty at the hour of the recording (between 13:35 and 14:05) with just two other players at a table nearby. Most of the snapshots from this scenario contain images of the pool game from various angles as Andrew walked around and leaned above the pool table, images of Andrew's friend, and images of the bar including furniture, tables, and the other two players; see Figure 8c for a few representative snapshots.

(4) *Traveling by train.* Andrew traveled by train to another city for an eye examination. The recording was made in one afternoon after 14:00. The distance traveled was about 30 km and the train stopped in four stations along the route. During this recording, Andrew was static, sitting down, and looking outside the window. The landscape was covered with a thin layer of snow as it was winter and the outside temperature was just slightly above zero degrees Celsius. Snapshots from this scenario include the railway tracks, various landscapes, and a few snapshots with people waiting in train stations; see Figure 8d.

(5) *Grocery shopping at a supermarket.* In this scenario, Andrew went to a supermarket. He used a large shopping cart and went through several departments, such as kitchen utensils, cleaning products and detergents, bathroom products, clothes, animal food, drinks, fruits and vegetables. Snapshots collected in this scenario comprise images of products, people shopping, and vendors; see Figure 8e. The location was indoor and well-lit. The recording ends before Andrew reaches the cash register.

(6) *Walking inside a building.* Andrew walked to several offices located in three buildings of the university (denoted by letters "A", "D", and "E") connected by corridors above the ground. Building A, the oldest one, had less well-lit corridors, building D had better natural light, and building E was the newest and the most modern building with an architectural style that favored sunlight to fill the rooms and corridors. Therefore, snapshots collected in this scenario have a variety of lighting conditions, representative of a wide range of indoor environments. Andrew walked down the hallways and staircases of the three buildings for 26 minutes and he stopped on a bench for 4 minutes. Figure 8f illustrates a few snapshots that were captured by the Life-Tags system in this scenario.

Scenario	N	Confidence ≥ 95%				Confidence ≥ 90%				Confidence ≥ 85%			
		Min	Max	M	SD	Min	Max	M	SD	Min	Max	M	SD
bar	3600	0	20	10.6	5.2	1	20	16.5	4.3	4	20	19.1	2.4
building	3600	0	20	6.3	3.2	1	20	12.8	4.4	3	20	17.5	3.6
bus	3600	0	20	8.5	4.5	2	20	16.1	3.9	5	20	19.4	1.7
restaurant	3600	0	20	6.4	3.1	2	20	12.0	4.0	3	20	17.0	3.3
supermarket	3600	0	20	6.5	4.6	1	20	12.3	5.1	2	20	16.8	3.9
train	3600	1	19	8.8	2.5	6	20	15.5	2.8	10	20	19.4	1.2
Total	**21,600**			**7.9**	**3.9**			**14.2**	**4.1**			**18.2**	**2.7**

Table 1. An overview of our data showing the average number of tags automatically extracted per snapshot for each scenario, according to various confidence levels. *Note:* in the rest of the paper, we focus on concepts recognized with at least 95% confidence; see the left part of the table.

5 RESULTS

We are interested for our technical evaluation in the effect of Sampling-Frequency and Motion-Threshold independent variables on the description richness of the lifelog in terms of (i) the number of distinct tags pertaining to visual concepts that are automatically extracted from snapshots (Num-Concepts), but also in terms of (ii) the size of the data (LifeLog-Size). In this section, we report on these effects in detail as we want to understand their impact on engineering Life-Tags with implications for any wearable video camera-based system for abstracting life with visual concepts. First, we present an overview of the data collected in our experiment.

5.1 Data Overview: Snapshots and Tags

We collected a total number of 21,600 snapshots corresponding to 180 minutes of recording time for all the six scenarios considered in our experiment. From these snapshots, the `Clarifai` cloud service extracted a total number of 169,435 tags with 95% confidence, of which 1,185 were distinct. Table 1 presents a summary of our data, showing the number of concepts that were extracted for each Scenario with confidence levels of at least 95%, 90%, and 85%, respectively. On average, 7.9 concepts (SD = 3.9) were extracted with 95% confidence or above, and 18.2 concepts (SD = 2.7) were extracted per snapshot with at least 85% confidence. For the rest of the paper, we focus just on the results pertaining to concepts with at least 95% confidence to minimize the influence of false positives; equivalently, the error rate of a false positive is less than 5%. The maximum number of concepts (with at least 95% confidence) was detected for the *bar* scenario (M = 10.6 tags per snapshot), while at the opposite end were *restaurant* (M = 6.4) and *supermarket* (M = 6.5), respectively; see Table 1. The average snapshot size was 353.2 KB (SD = 46.1 KB), varying between a minimum of 288.8 KB for the *building* scenario and a maximum of 424.2 KB for *train*; see Table 2 for descriptive statistics regarding the LifeLog-Size variable. The average compression rate was $(353.2 \cdot 1024\,\text{B})/(1920 \cdot 1080 \cdot 4\,\text{B}) = 4.4\%$ of the original size, or a 23:1 compression, corresponding to "medium quality" JPEG images according to the quality factor $Q = 25$. Storing all the N=21,600 snapshots for all the six scenarios considered in our experiment required 7.27 GB.

5.2 The effect of Sampling Frequency and Motion Detection variables

Table 3 illustrates the effect of Sampling-Frequency and Motion-Threshold, respectively, on the number of snapshots processed by Life-Tags for automatic tag extraction; *e.g.*, using a sampling frequency of one snapshot captured every 5 seconds, a total number of 360 snapshots were processed, while using a motion threshold of 30%, a total number of 2,571 images were processed on average for automatic tag extraction, all scenarios considered. At an average snapshot size of 353.2 Kb (see Table 2), it results that approximatively 886.8 MB of storage space is required per scenario. The

Scenario	N	Snapshot size† (KBytes) Min	Max	M	SD
bar	3600	164.8	444.8	362.1	36.1
building	3600	167.5	608.5	288.8	41.3
bus	3600	150.9	645.7	337.2	76.4
restaurant	3600	179.1	390.7	312.1	24.9
supermarket	3600	201.6	585.0	394.5	66.2
train	3600	279.6	521.5	424.2	31.5
Average		**190.6**	**532.7**	**353.2**	**46.1**

† For a compressed (JPEG, $Q = 25$), full HD (1920 × 1080 pixels) image

Table 2. An overview of our data listing descriptive statistics of the LifeLog-Size variable. On average, a compressed, full HD snapshot captured by Life-Tags required 353.2 KB to store.

Scenario	Sampling-Frequency 0.5	1	2	5	10	15	30	60	Motion-Threshold 10%	20%	30%	40%	50%
bar	3600	1800	900	360	180	120	60	30	3294	3046	2683	2180	1477
building	3600	1800	900	360	180	120	60	30	3102	2951	2662	2290	1819
bus	3600	1800	900	360	180	120	60	30	3192	2792	2163	1442	856
restaurant	3600	1800	900	360	180	120	60	30	3170	2850	2444	1992	1540
supermarket	3600	1800	900	360	180	120	60	30	3550	3471	3372	3242	2982
train	3600	1800	900	360	180	120	60	30	3567	3356	2099	677	141
Average	**3600**	**1800**	**900**	**360**	**180**	**120**	**60**	**30**	**3313**	**3078**	**2571**	**1971**	**1469**

Table 3. The number of snapshots considered for tag extraction, function of the Scenario, Sampling-Frequency, and Motion-Threshold independent variables. *Note:* the number of snapshots is constant for all scenarios for any given value of the Sampling-Frequency variable, *e.g.*, 1,800 images are stored when sampling one snapshot per second, and varies per scenario according to the Motion-Threshold level and the actual motion observed, *e.g.*, 2,951 snapshots were selected for automatic tag extraction in the *building* scenario when using a 20% threshold.

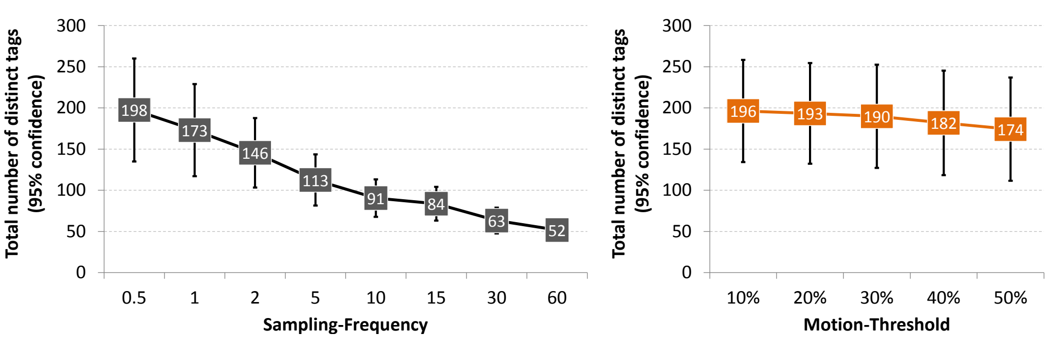

Fig. 9. The effect of Sampling-Frequency (left) and Motion-Threshold (right) on the total number of distinct concepts (Num-Concepts) automatically extracted from 30 minutes of recording, all scenarios considered. *Note:* error bars show 95% confidence intervals.

number of processed images ($N_p < N$) is inversely proportional to the sampling frequency for the uniform sampling approach, *i.e.*, $N_p = 1800 \cdot \text{Sampling-Frequency}^{-1}$ ($R^2 = 1.000$), while linearly related to the threshold used to detect motion between consecutive frames for the motion detection approach, *i.e.*, $N_p = -4795 \cdot \text{Motion-Threshold} + 3919$ ($R^2 = .982$), according to the figures from Table 3.

Scenario	Sampling-Frequency								Motion-Threshold				
	0.5	**1**	**2**	**5**	**10**	**15**	**30**	**60**	**10%**	**20%**	**30%**	**40%**	**50%**
bar	-	100%	100%	100%	100%	100%	100%	90%	100%	90%	80%	80%	80%
building	-	100%	100%	100%	100%	100%	90%	80%	100%	100%	90%	90%	90%
bus	-	100%	100%	100%	100%	90%	90%	90%	100%	90%	80%	70%	80%
restaurant	-	100%	100%	90%	100%	100%	100%	90%	100%	100%	100%	100%	100%
supermarket	-	100%	100%	100%	100%	90%	90%	70%	100%	100%	100%	100%	100%
train	-	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Average	-	**100%**	**100%**	**98.3%**	**100%**	**96.7%**	**95.0%**	**86.7%**	**100%**	**96.7%**	**91.7%**	**90.0%**	**91.7%**

Table 4. The percentage of tags still in the top-10 list when sampling snapshots at various frequency rates or when employing different thresholds for detecting motion. For example, for snapshots captured every 15 seconds, 96.7% of the tags are still present in the top-10 list. *Note:* Sampling-Frequency = 0.5 represents the control condition as it operates on the entire dataset, against which we compare all the other Sampling-Frequency and Motion-Threshold conditions.

	Sampling-Frequency							
	0.5	**1**	**2**	**5**	**10**	**15**	**30**	**60**
1	stock	stock	stock	stock	stock	stock	stock	stock
2	shopping	shopping	shopping	shopping	shopping	shopping	shopping	shopping
3	business	business	business	business	business	business	business	market
4	commerce	commerce	commerce	commerce	commerce	commerce	commerce	commerce
5	market	market	market	market	market	market	market	shop
6	shop	shop	shop	shop	shop	shop	shop	business
7	shelf	shelf	shelf	no person	industry	industry	industry	shelf
8	no person	no person	no person	industry	no person	no person	no person	sale
9	industry	industry	industry	shelf	city	shelf	shelf	people
10	city	city	city	city	shelf	people	people	supermarket

Table 5. Top-10 concepts for the *supermarket* scenario function of Sampling-Frequency.

Figure 9 illustrates the number of distinct concepts, Num-Concepts, function of the Sampling-Frequency and Motion-Threshold independent variables, respectively. The number of total distinct concepts decreased from 198 (average value, all scenarios considered) for one snapshot captured every half second to 52 for one snapshot captured every 60 seconds, representing a decrease of 74%; see Figure 9, left. There was also a decrease in Num-Concepts for higher Motion-Threshold values, yet less accentuated, *i.e.*, 11% from 196 to 174 distinct concepts corresponding to conditions 10% and 50%; see Figure 9, right. To understand better the influence of Sampling-Frequency and Motion-Threshold on the richness of the results in terms of the number of distinct concepts and their consistency along the various conditions of these two variables, we looked at the top-10 most frequent tags from each scenario; see Table 4 for a summary of the results. Overall, we found that the top-10 concepts were relatively stable, *i.e.*, sampling one snapshot every 60 seconds resulted in 86.7% of the concepts that were originally detected in the Sampling-Frequency = 0.5 condition to still be present in the top-10 list for Sampling-Frequency = 60; while using a motion threshold of 50% led to 91.7% of the original concepts to remain in the top-10 with slight variations depending on the scenario. To illustrate just one example, Tables 5 and 6 exemplify the influence of Sampling-Frequency and Motion-Threshold on the top-10 list of the concepts for *supermarket*.

5.3 Summary of Results

We found that the frequency at which snapshots are captured and the threshold employed to detect motion events have a practical impact on the number of snapshots that are effectively processed for tag extraction and, implicitly, on the number of distinct concepts that form the tag cloud for

	Motion-Threshold				
	10%	**20%**	**30%**	**40%**	**50%**
1	stock	stock	stock	stock	stock
2	shopping	shopping	shopping	shopping	shopping
3	business	business	business	commerce	commerce
4	commerce	commerce	commerce	business	business
5	market	market	market	market	market
6	shop	shop	shop	shop	shop
7	shelf	shelf	shelf	shelf	shelf
8	no person	industry	industry	industry	industry
9	industry	no person	city	city	city
10	city	city	no person	no person	no person

Table 6. Top-10 concepts for the *supermarket* scenario function of Motion-Threshold.

abstracting the lifelog. More snapshots per second and a lower threshold for motion detection resulted in a larger number of concepts, yet at the expense of more storage demands. While the effect of Sampling-Frequency was more accentuated (see Figure 9) with a decrease in the average number of distinct concepts from 198 to 52 between our extreme Sampling-Frequency conditions of 0.5 and 60, the effect of Motion-Threshold was relatively small, with just a 11% decrease for Num-Concepts for the extreme conditions corresponding to the 10% and 50% thresholds. In the following, we employ our Tag-Cloud-Efficiency measure, not reported so far, and the information from Figure 9 and Tables 2 and 3 to conclude on these results.

The abstracting Tag-Cloud-Efficiency, computed as the ratio between the number of distinct concepts that form the tag cloud for abstracting life and the actual memory needed to store the corresponding snapshots from which tags were extracted, varies for uniform sampling from 198 / (3600 · 353.2 KB / 1024) = 0.159 concepts per MB when sampling two snapshots every second to 52 / (30 · 353.2 KB / 1024) = 5.025 when sampling one snapshot every 60 seconds; see the values listed in Figure 9 and Table 2. For the motion detection approach, Tag-Cloud-Efficiency varied from 196 / (3313 · 353.2 KB / 1024) = 0.172 distinct concepts per MB when detecting motion with a 10% threshold to 174 / (1469 · 353.2 KB / 1024) = 0.343 concepts per MB for a 50% threshold; see the values in Figure 9 and Table 3. These results show that recording life by taking snapshots based on event detection, such as motion, appears to be reasonably efficient in terms of both Num-Concepts and LifeLog-Size. Even in the most extreme condition with a 50% motion detection threshold, the number of distinct concepts is just 12% smaller than when processing all the snapshots (*i.e.*, 174 *vs.* 198 concepts), while using a threshold of 30% to detect motion events results in an average number of distinct concepts that is just 4% smaller than in the control condition (*i.e.*, 190 *vs.* 198 concepts); see Figure 9. While the Tag-Cloud-Efficiency measure is maximized for uniform sampling of one snapshot every 60 seconds, the number of distinct concepts drops severely, from 198 to 52. Based on these findings, our recommendation is to sample snapshots based on events, such as motion, which will result in a high number of distinct visual concepts to form the tag cloud, while providing a good compromise in terms of the memory required to store the snapshots; see Table 3.

6 CONCLUSION AND FUTURE WORK

We presented Life-Tags, a system for abstracting life by generating tag clouds of concepts reflective of the user's visual experience. We focused on the design criteria for Life-Tags and on engineering aspects to ensure a rich tag cloud of concepts, comprehensive of the life experienced by the lifelogger. Interesting future work will address integration of other sensors and input devices with the Life-Tags system, such as mobile eye trackers [32] or smart rings [20], inter-connected by flexible event-based software architecture [48], as well as employing models of visual attention [38] to generate and present users with a variety of tag clouds containing selected lists of concepts, such as concepts pertaining to scenes likely not attended by the user's visual system at the moment of the recording. Also, future work will focus on evaluating usability aspects of Life-Tags with participants of diverse categories of gender, age, and professional occupation as well as studying the effect of tag clouds to support memory for practical eHealth and mHealth applications. We hope that our contributions and empirical results will inspire the Engineering Interactive Computing Systems community to explore new techniques and prototypes to *abstract life* using wearable sensors and devices towards generating, presenting, and sharing informative and comprehensive digital summaries of our life experiences.

ACKNOWLEDGMENTS

This work was supported by a grant of the Ministry of Research and Innovation, CNCS-UEFISCDI, project no. PN-III-P1-1.1-TE-2016-2173 (TE141/2018), within PNCDI III. Original versions of the icons used in Figure 3 were made by Freepik (http://www.freepik.com, "Miscellaneous Elements" pack) from Flaticon (http://www.flaticon.com), licensed under Creative Commons BY 3.0 (http://creativecommons.org/licenses/by/3.0).

REFERENCES

[1] R. Agarwal, N. Ladha, M. Agarwal, K. K. Majee, A. Das, S. Kumar, S. K. Rai, A. K. Singh, S. Nayak, S. Dey, R. Dey, and H. N. Saha. 2017. Low cost ultrasonic smart glasses for blind. In *Proceedings of the 2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. 210–213. https://doi.org/10.1109/IEMCON.2017.8117194

[2] Adrian Aiordăchioae, Radu-Daniel Vatavu, and Dorin-Mircea Popovici. 2019. A Design Space for Vehicular LifeLogging to Support Creation of Digital Content in Connected Cars. In *Proceedings of the 11th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '19)*. https://doi.org/10.1145/3319499.3328234

[3] Abdullah Al Mahmud, Jeffrey Braun, and Jean-Bernard Martens. 2010. Designing to Capture and Share Life Experiences for Persons with Aphasia. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '10)*. ACM, New York, NY, USA, 391–392. https://doi.org/10.1145/1851600.1851680

[4] Amazon. 2019. Amazon Rekognition - Video and Image - AWS. Retrieved April 15, 2019 from https://aws.amazon.com/rekognition/

[5] Olivier Augereau, Charles Lima Sanches, Koichi Kise, and Kai Kunze. 2018. Wordometer Systems for Everyday Life. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 123 (Jan. 2018), 21 pages. https://doi.org/10.1145/3161601

[6] AYLIEN. 2019. Image Tagging | AYLIEN. Retrieved April 15, 2019 from https://aylien.com/text-api/image-tagging/

[7] Michael Barney, Jonathan Kilner, Gilmar Brito, Aida Araújo, and Meuse Nogueira. 2017. Sensory Glasses for the Visually Impaired. In *Proceedings of the 14th Web for All Conference on The Future of Accessible Work (W4A '17)*. ACM, New York, NY, USA, Article 27, 2 pages. https://doi.org/10.1145/3058555.3058584

[8] E. Berry, N. Kapur, L. Williams, S. Hodges, P. Watson, G. Smyth, J. Srinivasan, R. Smith, B. Wilson, and K. Wood. 2007. The use of a wearable camera, SenseCam, as a pictorial diary to improve autobiographical memory in a patient with limbic encephalitis: A preliminary report. *Neuropsychol. Rehabil.* 17, 4–5 (Aug-Oct 2007), 582–601. https://doi.org/10.1080/09602010601029780

[9] Mark Blum, Alex (Sandy) Pentland, and Gerhard Troster. 2006. InSense: Interest-Based Life Logging. *IEEE MultiMedia* 13, 4 (Oct. 2006), 40–48. https://doi.org/10.1109/MMUL.2006.87

[10] Tianlang Chen, Yuxiao Chen, and Jiebo Luo. 2017. A Selfie is Worth a Thousand Words: Mining Personal Patterns Behind User Selfie-posting Behaviours. In *Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 23–31. https://doi.org/10.1145/3041021.3054142

[11] Soumyadeb Chowdhury, Md Sadek Ferdous, and Joemon M Jose. 2016. Bystander Privacy in Lifelogging. In *Proceedings of the 30th International BCS Human Computer Interaction Conference: Companion Volume (HCI '16)*. BCS Learning & Development Ltd., Swindon, UK, Article 15, 3 pages. https://doi.org/10.14236/ewic/HCI2016.62

[12] Clarifai. 2019. About Clarifai. Transforming how we see the world. Retrieved January 25, 2019 from https://clarifai.com/about

[13] Google Clips. 2019. Google Clips Specifications. Retrieved February 12, 2019 from https://support.google.com/googleclips/answer/7545447?hl=en

[14] Will Coldwell. 2014. Autographer wearable camera: reviewed on a Berlin city break. Retrieved February 12, 2019 from https://www.theguardian.com/travel/2014/feb/27/autographer-wearable-camera-review-berlin-city-break

[15] Dirk de Jager, Alex L. Wood, Geoff V. Merrett, Bashir M. Al-Hashimi, Kieron O'Hara, Nigel R. Shadbolt, and Wendy Hall. 2011. A Low-power, Distributed, Pervasive Healthcare System for Supporting Memory. In *Proceedings of the First ACM MobiHoc Workshop on Pervasive Wireless Healthcare (MobileHealth '11)*. ACM, New York, NY, USA, Article 5, 7 pages. https://doi.org/10.1145/2007036.2007043

[16] Yuan Ding, Yuan Du, Yingkai Hu, Zhengye Liu, Luqin Wang, Keith Ross, and Anindya Ghose. 2011. Broadcast Yourself: Understanding YouTube Uploaders. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference (IMC '11)*. ACM, New York, NY, USA, 361–370. https://doi.org/10.1145/2068816.2068850

[17] Martin Dodge and Rob Kitchin. 2007. 'Outlines of a World Coming into Existence': Pervasive Computing and the Ethics of Forgetting. *Environment and Planning B: Planning and Design* 34, 3 (2007), 431–445. https://doi.org/10.1068/b32041t

[18] Gartner. 2018. Gartner says worldwide wearable device sales to grow 26 percent in 2019. Retrieved February 12, 2019 from https://www.gartner.com/en/newsroom/press-releases/2018-11-29-gartner-says-worldwide-wearable-device-sales-to-grow-

[19] Jim Gemmell. 2014. Life-logging, Thing-logging and the Internet of Things. In *Proceedings of the 2014 Workshop on Physical Analytics (WPA '14)*. ACM, New York, NY, USA, 17–17. https://doi.org/10.1145/2611264.2611276

[20] Bogdan-Florin Gheran, Jean Vanderdonckt, and Radu-Daniel Vatavu. 2018. Gestures for Smart Rings: Empirical Results, Insights, and Design Implications. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 623–635. https://doi.org/10.1145/3196709.3196741

[21] Google. 2019. Vision AI. Derive Image Insights with ML. Google Cloud. Retrieved April 15, 2019 from https://cloud.google.com/vision/

[22] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. 2014. LifeLogging: Personal Big Data. *Found. Trends Inf. Retr.* 8, 1 (June 2014), 1–125. https://doi.org/10.1561/1500000033

[23] Phil Hall. 2019. Best action camera 2019: 10 cameras for the GoPro generation. Retrieved February 12, 2019 from https://www.techradar.com/news/best-action-camera

[24] Hugh Hart. 2010. April 14, 1996: JenniCam starts lifecasting. Retrieved April 14, 2019 from https://www.wired.com/2010/04/0414jennicam-launches/

[25] Julia Heidemann, Mathias Klier, and Florian Probst. 2012. Online social networks: A survey of a global phenomenon. *Computer Networks* 56, 18 (2012), 3866–3878. https://doi.org/10.1016/j.comnet.2012.08.009

[26] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. 2006. SenseCam: A Retrospective Memory Aid. In *Proceedings of the 8th International Conference on Ubiquitous Computing (UbiComp'06)*. Springer-Verlag, Berlin, Heidelberg, 177–193. https://doi.org/10.1007/11853565_11

[27] Roberto Hoyle, Robert Templeman, Steven Armes, Denise Anthony, David Crandall, and Apu Kapadia. 2014. Privacy Behaviors of Lifeloggers Using Wearable Cameras. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, New York, NY, USA, 571–582. https://doi.org/10.1145/2632048.2632079

[28] Chunjia Hu, Guangtao Zhai, and Duo Li. 2015. An Augmented-Reality night vision enhancement application for see-through glasses. In *2015 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*. 1–6. https://doi.org/10.1109/ICMEW.2015.7169860

[29] Rui-Ting Huang. 2018. What Motivates People to Continuously Post Selfies? The Moderating Role of Perceived Relative Advantage. *Comput. Hum. Behav.* 80, C (March 2018), 103–111. https://doi.org/10.1016/j.chb.2017.11.007

[30] Mohammad Hossein Jarrahi, Nicci Gafinowitz, and Grace Shin. 2018. Activity Trackers, Prior Motivation, and Perceived Informational and Motivational Affordances. *Personal Ubiquitous Comput.* 22, 2 (April 2018), 433–448. https://doi.org/10.1007/s00779-017-1099-9

[31] Hayeon Jeong, Heepyung Kim, Rihun Kim, Uichin Lee, and Yong Jeong. 2017. Smartwatch Wearing Behavior Analysis: A Longitudinal Study. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 60 (Sept. 2017), 31 pages. https://doi.org/10.1145/3131892

[32] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 1151–1160. https://doi.org/10.1145/2638728.2641695

[33] Keigo Kitamura, Toshihiko Yamasaki, and Kiyoharu Aizawa. 2008. Food Log by Analyzing Food Images. In *Proceedings of the 16th ACM International Conference on Multimedia (MM '08)*. ACM, New York, NY, USA, 999–1000. https://doi.org/10.1145/1459359.1459548

[34] Marion Koelle, Matthias Kranz, and Andreas Möller. 2015. Don't Look at Me That Way!: Understanding User Attitudes Towards Data Glasses Usage. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. ACM, New York, NY, USA, 362–372. https://doi.org/10.1145/2785830.2785842

[35] Hyowon Lee, Alan F. Smeaton, Noel E. O'connor, Gareth Jones, Michael Blighe, Daragh Byrne, Aiden Doherty, and Cathal Gurrin. 2008. Constructing a SenseCam Visual Diary As a Media Process. *Multimedia Syst.* 14, 6 (Dec. 2008), 341–349. https://doi.org/10.1007/s00530-008-0129-x

[36] Andrés Lucero. 2018. Living Without a Mobile Phone: An Autoethnography. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 765–776. https://doi.org/10.1145/3196709.3196731

[37] J. Machajdik, A. Hanbury, A. Garz, and R. Sablatnig. 2011. Affective Computing for Wearable Diary and Lifelogging Systems: An Overview. In *Proceedings of the 35th Workshop of the Austrian Association for Pattern Recognition*. Austrian Computer Society. https://www.ec.tuwien.ac.at/node/6421

[38] Matei Mancas, Vincent P. Ferrera, Nicolas Riche, and John G. Taylor. 2016. *From Human Attention to Computational Attention: A Multidisciplinary Approach* (1st ed.). Springer Publishing Company, Inc. https://doi.org/10.1007/978-1-4939-3435-5

[39] Steve Mann, James Fung, and Eric Moncrieff. 1999. EyeTap Technology for Wireless Electronic News Gathering. *SIGMOBILE Mob. Comput. Commun. Rev.* 3, 4 (Oct. 1999), 19–26. https://doi.org/10.1145/584039.584044

[40] Joshua McVeigh-Schultz, Jennifer Stein, Jacob Boyle, Emily Duff, Jeff Watson, Avimaan Syam, Amanda Tasse, Simon Wiscombe, and Scott Fisher. 2012. Vehicular Lifelogging: New Contexts and Methodologies for Human-car Interaction. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*. ACM, New York, NY, USA, 221–230. https://doi.org/10.1145/2212776.2212800

[41] MeCam. 2019. High definition video camera | Best life logging device | Mini video camera - MeCam. Retrieved February 12, 2019 from https://mecam.me/products/mecam-hd

[42] Vangelis Metsis, Dimitrios Kosmopoulos, Vassilis Athitsos, and Fillia Makedon. 2014. Non-invasive Analysis of Sleep Patterns via Multimodal Sensor Input. *Personal and Ubiquitous Computing* 18, 1 (Jan. 2014), 19–26. https://doi.org/10.1007/s00779-012-0623-1

[43] Narrative. 2019. The World's Most Wearable HD Video Camera | Narrative Clip 2. Retrieved February 12, 2019 from http://getnarrative.com/

[44] Donald A. Norman. 2004. *Emotional Design. Why We Love (or Hate) Everyday Things.* Basic Books, New York, NY, USA. https://www.nngroup.com/books/emotional-design/

[45] Bogdan Pogorelc, Artur Lugmayr, Bjorn Stockleben, Radu-Daniel Vatavu, Nina Tahmasebi, Estefania Serral, Emilija Stojmenova, Bojan Imperl, Thomas Risse, Gideon Zenz, and Matjaz Gams. 2013. Ambient Bloom: New Business, Content, Design and Models to Increase the Semantic Ambient Media Experience. *Multimedia Tools and Applications* 66, 1 (2013), 7–32. https://doi.org/10.1007/s11042-012-1228-4

[46] Andrei Popleteev, Nicolas Louveton, and Roderick McCall. 2015. Colorizer: Smart Glasses Aid for the Colorblind. In *Proceedings of the 2015 Workshop on Wearable Systems and Applications (WearSys '15)*. ACM, New York, NY, USA, 7–8. https://doi.org/10.1145/2753509.2753516

[47] RedShed. 2018. The 8 Best Spy Glasses of 2018. Retrieved January 25, 2019 from https://redshed.co.uk/tech-reviews/best-spy-glasses/

[48] Ovidiu-Andrei Schipor, Radu-Daniel Vatavu, and Jean Vanderdonckt. 2019. Euphoria: A Scalable, Event-Driven Architecture for Designing Interactions Across Heterogeneous Devices in Smart Environments. *Information and Software Technology* 109 (2019), 43–59. https://doi.org/10.1016/j.infsof.2019.01.006

[49] Courtney Seiter. 2016. The Psychology of Social Media: Why We Like, Comment, and Share Online. Retrieved January 25, 2019 from https://blog.bufferapp.com/psychology-of-social-media

[50] Abigail J. Sellen and Steve Whittaker. 2010. Beyond Total Capture: A Constructive Critique of Lifelogging. *Commun. ACM* 53, 5 (May 2010), 70–77. https://doi.org/10.1145/1735223.1735243

[51] Kristen Shinohara and Josh Tenenberg. 2007. Observing Sara: A Case Study of a Blind Person's Interactions with Technology. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (Assets '07)*. ACM, New York, NY, USA, 171–178. https://doi.org/10.1145/1296843.1296873

[52] Snap. 2019. Spectacles by Snapchat | Your Hands-Free Camera. Retrieved January 25, 2019 from https://www.spectacles.com/

[53] SnapCam. 2019. iON USA | SnapCam. Retrieved February 12, 2019 from https://usa.ioncamera.com/snapcam/

[54] Flávio Souza, Diego de Las Casas, Vinícius Flores, SunBum Youn, Meeyoung Cha, Daniele Quercia, and Virgílio Almeida. 2015. Dawn of the Selfie Era: The Whos, Wheres, and Hows of Selfies on Instagram. In *Proceedings of the 2015 ACM on Conference on Online Social Networks (COSN '15)*. ACM, New York, NY, USA, 221–231. https://doi.org/10.1145/2817946.2817948

[55] Gary Stix. 2011. Photographic Memory: Wearable Cam Could Help Patients Stave Off Effects of Impaired Recall. *Scientific American* (October 2011). https://www.scientificamerican.com/article/photographic-memory-wearable/

[56] Radu-Daniel Vatavu. 2012. Presence bubbles: supporting and enhancing human-human interaction with ambient media. *Multimedia Tools and Applications* 58, 2 (May 2012), 371–383. https://doi.org/10.1007/s11042-010-0674-0

[57] Seung-Ho Yang, Hyun-Woo Kim, and Min Young Kim. 2011. Human Visual Augmentation Using Wearable Glasses with Multiple Cameras and Information Fusion of Human Eye Tracking and Scene Understanding. In *Proceedings of the 6th International Conference on Human-robot Interaction (HRI '11)*. ACM, New York, NY, USA, 287–288. https://doi.org/10.1145/1957656.1957774

[58] Hinbarji Z., Albatal R., O'Connor N., and Gurrin C. 2016. *LoggerMan, a Comprehensive Logging and Visualization Tool to Capture Computer Usage*. Vol. 9517. Springer, Cham. https://doi.org/10.1007/978-3-319-27674-8_31

[59] Yuhang Zhao, Sarit Szpiro, and Shiri Azenkot. 2015. ForeSee: A Customizable Head-Mounted Vision Enhancement System for People with Low Vision. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '15)*. ACM, New York, NY, USA, 239–249. https://doi.org/10.1145/2700648.2809865

[60] Yuhang Zhao, Sarit Szpiro, Jonathan Knighten, and Shiri Azenkot. 2016. CueSee: Exploring Visual Cues for People with Low Vision to Facilitate a Visual Search Task. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 73–84. https://doi.org/10.1145/2971648.2971730

[61] Liting Zhou, Zaher Hinbarji, Duc-Tien Dang-Nguyen, and Cathal Gurrin. 2018. LIFER: An Interactive Lifelog Retrieval System. In *Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge (LSC '18)*. ACM, New York, NY, USA, 9–14. https://doi.org/10.1145/3210539.3210542

[62] Floriano Zini, Martin Reinstadler, and Francesco Ricci. 2015. Life-logs Aggregation for Quality of Life Monitoring. In *Proceedings of the 5th International Conference on Digital Health 2015 (DH '15)*. ACM, New York, NY, USA, 131–132. https://doi.org/10.1145/2750511.2750531

Received February 2019; revised March 2019; accepted April 2019