Multimedia Memory Cues for Augmenting Human Memory

Tilman Dingler, Passant El Agroudy, Huy Viet Le, and Albrecht Schmidt
University of Stuttgart, Germany

Evangelos Niforatos, Agon Bexheti, and Marc Langheinrich
Università della Svizzera Italiana (USI), Lugano, Switzerland

Human memory has long been used as an important tool in helping people effectively perform daily tasks. We write down information that we don't want to forget, or tie a knot into a handkerchief to remember an important event. Today's technology offers many replacements for these tried and tested tools, such as electronic phone books, diaries with automated alarms, and even location-based reminders. Lifelogging—"a phenomenon whereby people can digitally record their own daily lives in varying amounts of detail"[1]—offers a powerful new set of tools to augment our memories.

In particular, the prospect of capturing a continuous stream of images or videos from both a first-person perspective and various third-person perspectives promises an unprecedented level of rich multimedia content. Such content could disclose a significant amount of detail, given the right set of analysis tools. Having comprehensive recordings of our lives would make it possible, at least in principle, to search such an electronic diary for any kind of information that might have been forgotten or simply overlooked: "What was the name of the new colleague that I met yesterday?" or "Where did I last see my keys?"

In the context of the EU-funded *Recall* project (http://recall-fet.eu), we also look into the use of such multimedia data to augment human memory—but in a conceptually different fashion. Instead of seeking to offer users an index that can be searched at any time, thereby diminishing the importance of their own memory, we seek to create a system that will measurably improve each user's own memory. Instead of asking yourself (that is, your electronic diary) for the name of the new colleague during your next encounter (which could be awkward as you wait for the diary to pull up the name), Recall users would have already trained their own memory to simply remember the colleague's name.

Here, we present the core research ideas of Recall, outlining the particular challenges of such an approach for multimedia research and summarizing the project's initial results. Our overall approach is to collect multimedia lifelog data and contextual information through a range of capture devices, process the captured data to create appropriate memory cues for later playback, and apply theories from psychology to develop tools and applications for memory augmentation (see Table 1).

Memory Cues

A system that aims to improve the user's own memory must be able to properly select, process, and present "memory cues." A memory cue is simply something that helps us remember—it is a snippet of information that helps us access a memory.[2] Figure 1 gives an overview of contextual information sources that produce these cues. Almost anything can work as a memory cue: a piece of driftwood might remind us of family vacations at the beach, an old song might remind us of our first high school dance, or the smell of beeswax might remind us of a childhood Christmas.

Multimedia—audio, pictures, video, and so on—is thus of particular interest. It holds a significant amount of information that can offer rich triggers for memory recollection. Furthermore, given today's technology, multimedia memory cues are relatively easy to capture. Recall uses memory cues to stimulate pathways in a user's memory that will reinforce the ability to retrieve certain information when needed in the future.

To be useful, memory cues thus don't have to actually contain all of the information needed. For example, a picture of a particular whiteboard drawing might not be detailed enough to show the individual labels, yet seeing a picture of the overall situation might be enough for a user to vividly remember not only

1070-986X/16/$31.00 © 2016 IEEE

Table 1. Summary of Recall studies to identify efficient cues for triggering memories.

System layer	Research probe	Data-capture approach	Recall-supporting cues
Capture memory cues	**On-body camera position:** How does the position affect the quality and perception of captured photos?	Automatic fixed-interval capture	Videos and photos from cameras (head-worn cameras offer better autobiographical cues, while chest-worn cameras are more stable) Faces (most relevant cues)
	PulseCam app: How can we capture only important photos using biophysical data?	Pulse-rate-triggered capture	Smaller number of captured photos for important activities
	MGOK app: How can we enhance the quality of memory cues by capturing more significant moments?	Limit number of pictures per day	Smaller number of captured photos for important activities
Extract memory cues	**Summarization of lifelog image collection:** What are the guidelines to produce video summaries from such collections?	Automatic fixed-interval capture	Summarized-video requirements: no more than three minutes; include people, objects, or actions; and present in the same chronological order
	Summarization of desktop activity screenshots: How can we reduce the volume of screenshots without affecting recall quality?	Reading-triggered capture (using a commercial eye tracker)	Smaller number of captured screenshots for important activities
	LISA prototype: How can we create a holistic and interactive solution for reflecting on daily activities?	Auto-sync with third-party services	Aggregated dashboard of projected visualizations and speech (location, pictures, fitness data, and calendar events)
Present memory cues	**EmoSnaps app**: How can we enhance emotional recall of past experiences using visual cues?	Capture at predefined moments (such as when a device is unlocked)	Selfies (facial expressions)
	Re-Live the Moment app: How can we use personalized multimedia cues to foster positive behavioral change in running?	Running-triggered capture using a music playlist (continuous capturing) and route photos (automatic, fixed-interval capture)	Time-lapse video: route-captured photos and personal running music playlist
	Déjà vu concept: How can we exploit priming to display ambient information about future situations to make them familiar?	Search for relevant information in third-party services	Visualizations of proactive information chunks about future situations

the diagram itself but also the discussion surrounding its creation. Similarly, an image with the face of a new colleague, together with the first letter of her name, might be enough to trigger our own recollection of the full name, and thus priming us to retrieve the full name when we meet the new colleague again.

Memory Capture

Near-continuous collection of memory cues (lifelogging) has become possible through a number of available technologies. Lifelog cameras, such as Microsoft's SenseCam (http://research.microsoft.com/en-us/projects/sensecam) or the Narrative Clip (http://getnarrative.com), let users capture the day in images. Every 10 to 120 seconds, these devices take a picture, culminating in a time-lapse sequence of images that can span days, weeks, or months. Additional audio and video footage can be collected through other user-worn capture devices and through cameras and microphones placed in the environment. All this data makes up *lifelogs*, but the quality of the footage often is volatile.

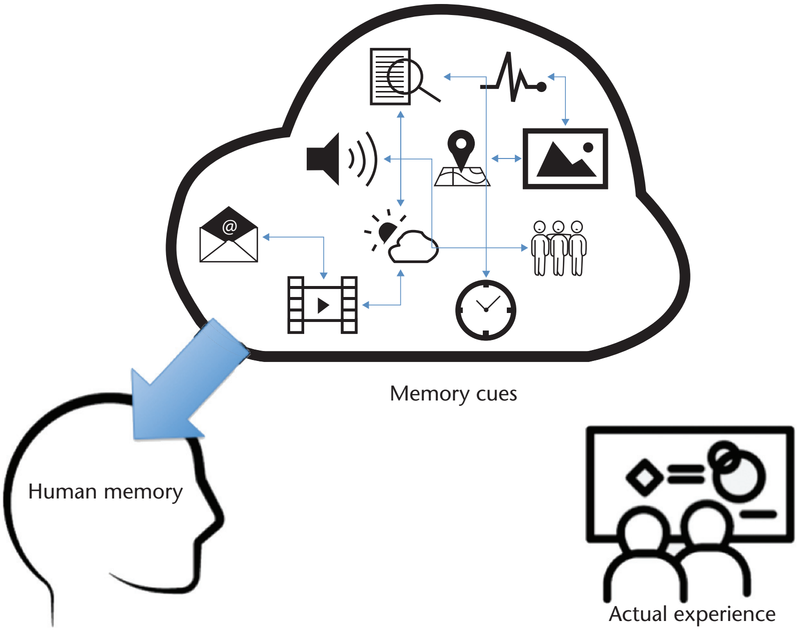


Figure 1. Contextual information sources that produce cues for supporting one's ability to recall a past experience of a future event.

Here, device positioning matters. We compared the body position where such cameras are normally attached—head or chest—and its effect on image quality and user perception.[3] We equipped 30 participants with cameras on their foreheads and chests and later asked them about their perception of the images collected. Additionally, we applied a set of standard image-processing algorithms to classify images, including sharpness filters and face and hand detection.

We learned that the chest-worn devices produced more stable and less motion-blurred images, through which feature detection by image processing algorithms worked better. Head-worn video cameras, on the other hand, captured more important autobiographical cues than chest-worn devices. Here, faces were shown to be most relevant for recall.

Beyond visual data, there is a number of different data types capable of enhancing lifelogs. Combining different data streams can form a more comprehensive picture: accelerometers, blood pressure, or galvanic skin response sensors, for example, output physiological data that can be used to assess the significance of images taken. Smartphones and watches often have some of these sensors already included. They further allow the collection of context information, such as time, location, or activities.

We proposed using biophysical data to distinguish between highly important and rather irrelevant moments, subsequently driving image capture. As such, we developed the PulseCam (see Figure 2a), an Android Wear and mobile app that takes the user's pulse rate to capture images of greater importance.[4] Eventually, merged with third-party data sources—such as calendar entries, email communication, or social network activities—lifelogs can be enriched with a holistic picture of a person's on- and offline activities throughout the day.

Most of this data can be collected implicitly—that is, without the user having to manually trigger the recording (by taking a picture, for example). Explicit recording, on the other hand, would imply the conscious act of recording a memory, as would a manual posting on

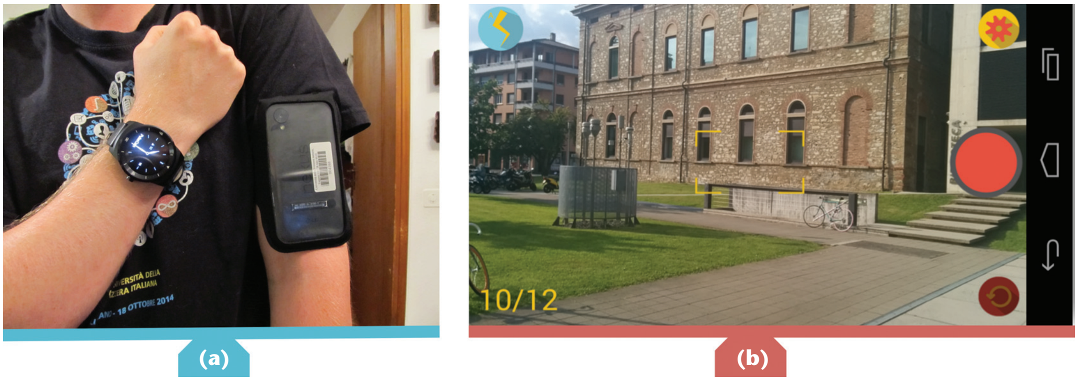


Figure 2. Screenshots of Recall probes for capturing memory cues. (a) The PulseCam prototype hardware. On the left arm, the user has an LG G smart watch for continuous heart rate capture. On the right arm, a Nexus S smartphone is attached to capture the pictures. (b) A screenshot of the My Good Old Kodak application (https://play.google.com/store/apps/details?id=ch.usi.inf.recall.myoldkodak&hl=en). The number of remaining photos in the day is displayed in the lower left corner.

Twitter, Facebook, or Instagram. Explicit capture tends to indicate the significance of a particular moment for a person, whereas implicit capture helps us ensure we don't miss key events. A combination of both capture modes leads to a richer lifelog, from which more significant memory cues might be drawn. To produce a holistic record, data coming from all these different sources must be properly time-synchronized, unless additional algorithms can infer temporal co-location from overlapping information (for example, a smartphone camera and chest-mounted camera showing two different viewpoints of the same scene) and thus post-hoc synchronize two or more streams.

To contrast how implicit and explicit captures influence our original (uncued) recollection of an event, and how well they can serve as memory cues, we developed the My Good Old Kodak (MGOK) mobile app. MGOK is a mobile camera application that artificially limits the amount of pictures that can be taken, resembling the classic film cameras (see Figure 2b).[5] We are currently analyzing data from a large trial that we ran with almost 100 students, snapping away for a day with a chest-worn lifelogging camera (implicit, unbounded), a "normal" smartphone camera app (explicit, unbounded), our MGOK app (explicit, bounded), or no camera at all. We hypothesize that the imposed capture limitation will result in moments of higher significance being captured, potentially leading to pictures that better serve as memory cues.

Memory Cue Extraction

Memory cues are stimuli hints that trigger the recall of a past experience or future event.[2] Cues can be presented, for example, on peripheral displays throughout the user's home or on a personal device, such as a smartphone. They are meant to trigger *episodic* (remembering past events) and *prospective* (remembering planned future events) memory recall. By frequently encountering certain cues, they can improve people's ability to recall a relevant memory and its details over time. Such an effect could potentially persist without the need for further technological support.

Thus, we currently focus on two main directions: summarizing large datasets of images, and merging and summarizing heterogeneous data resources. One of the main objectives is to find high-quality cues for efficiently triggering memories and recall. Compared to traditional summaries, an effective memory cue is minimalistic by itself but allows a wide range of associations to be made. It acts as a trigger to your own memory with all the richness that the memory entails.

Summarizing Large Datasets

Images, videos, and speech streams are a rich pool of information, because they capture experiences along with contextual cues such as locations or emotions in great detail. However, they require significant time to be moderated and viewed, creating the need for efficient automatic summarization techniques. For example, a single Narrative Clip that takes a picture every 30 seconds will produce approximately 1,500 pictures per day. On the other hand, the variety of digital services that we use on a daily basis produces a heterogeneous pool of data. This makes it hard to gain deeper insights or derive more general patterns. By merging sensor data and extracting meaning, we can derive holistic and meaningful insights.

Summarizing a large image collection. To inform and automatically generate lifelog summaries, we conducted a set of user studies to elicit design guidelines for video summaries.[6] We instructed 16 participants to create video summaries from their own lifelogging images and compared the results to nonsummary review techniques, such as using time lapses and reviews through an image browser. The three techniques were equally effective, but participants preferred the experience of their own video summaries.

However, such manual processing isn't always possible, especially when considering the large amounts of data collected in just one day. Insights from the preceding study lead us to the following set of guidelines, which we used to build a system for automating the creation of video summaries.

First, video summaries should not exceed three minutes, because most users don't want to spend an exhaustive amount of time reviewing lifelogging activities.

Second, images featuring combinations of people, places, objects, or actions are reportedly the most effective memory cues. These can be further enhanced by adding metadata, which improves the user's understanding of the image's context.

Finally, presenting images in a chronological order provides additional support (chronological, contextual, and inferential) for the reconstruction of memories. In particular, this affects

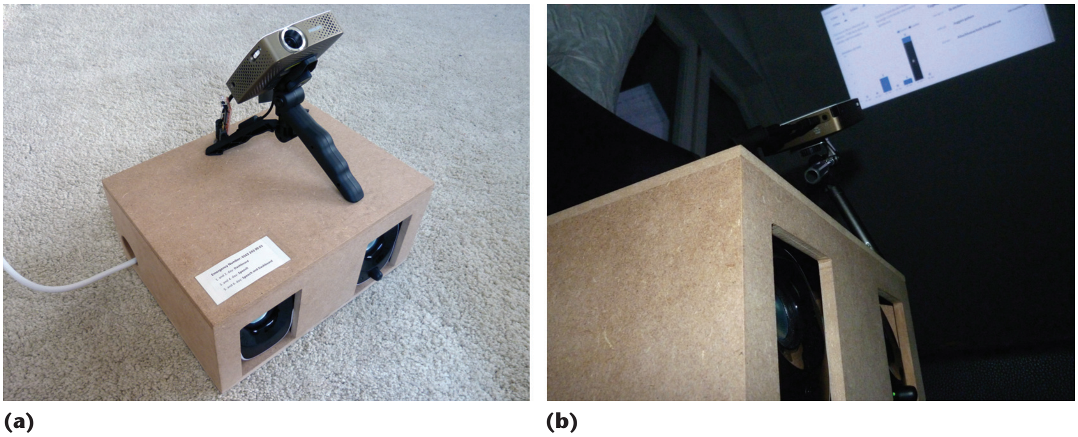


Figure 3. Pictures of the LISA prototype used to extract memory cues: (a) The prototype hardware, composed of a projector and a set of speakers. (b) The summary of memory cues presented as a dashboard projection and audio summary upon waking up.

activities with greater movement (such as sports activities, walking, or social events) because such activities require multiple images to cover additional details not well represented by a single image.

Summarizing desktop screenshots. To create a holistic lifelog of people's daily activities, we also need to look at people's technology usage across the day. Therefore, we investigated how to capture people's PC usage, as represented by their activities on their computer desktop. We focused on automatic screenshots, which—when triggered in a regular time interval—produce a large amount of images. This led to an investigation of how to minimize the sheer volume of snapshots taken by an automatic desktop logger in a work environment.[7] We compared three triggers for such snapshots: a fixed-time interval (two minutes) and two techniques informed by eye-tracking data—whenever the user's eye gaze focused on an application window, or whenever a reading activity was registered. Reading detection turned out to significantly reduce the amount of images taken while still capturing relevant activities.

Merging and Summarizing Heterogeneous Data Sources

There is a wide range of personal data streams accessible not only through capture devices but also through Web APIs and interaction logs. We created a projection system called Life Intelligence Software Assistant (LISA) in the form of a bedside device that provides a morning briefing, combining data from the past day with upcoming events (see Figure 3). Using visual projection and speech, it presents information from different data sources: locations visited, fitness stats, images taken, and calendar events.

In a pilot study, we found a mixture of speech and projection to be preferable to either of them alone. In a series of domestic deployments, we are currently investigating the effectiveness of different cues, display locations, and use cases.

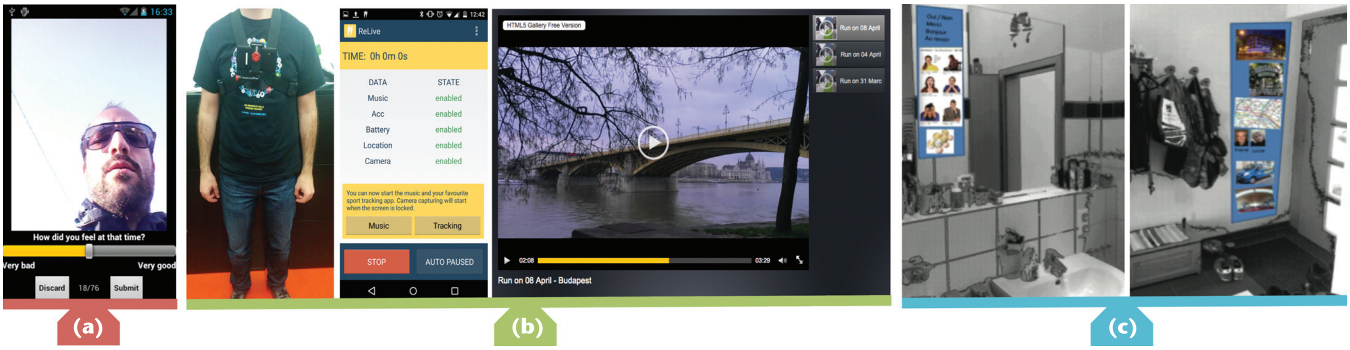
Presentation: Apps and Concepts

The capture of effective memory cues is essential to enable recall. A single effective cue can produce a great amount of details in memory, such that a comprehensive media capture of that same memory becomes redundant. To investigate the efficiency of certain memory cues, we created and deployed a series of presentation prototypes that allowed us to test how replaying captured cues would actually help participants remember prior experiences.

EmoSnaps

Initially, we investigated how visual cues can enhance emotional recall in the form of selfies. As such, we developed a mobile app called EmoSnaps (see Figure 4a). It unobtrusively captures pictures of the user's facial expression at predefined moments (such as when the user unlocks his or her smartphone).[8] Participants correctly identified the emotion captured on their selfies during a past moment solely by revisiting the selfie taken.

Surprisingly, participants managed to identify older pictures better than newer ones. We



You can now start the music and your favourite sport tracking app. Camera capturing will start when the screen is locked.

Figure 4. Screen shots of Recall research problems for presenting memory cues. (a) EmoSnaps—the user is asked to recall his or her emotion using a past selfie (https://play.google.com/store/apps/details?id=com.nifo.emosnaps&hl=en). (b) Re-Live the Moment—the user wears a chest-mounted smarphone. He starts running with his favorite music and tracking app. After the run, the user can check his "personal music clip." (c) Déjà vu—a conceptual prototype of context-aware peripheral displays showing information about relevant people and locations.

attribute this phenomenon to the probable conflict between recognizing emotion through facial expressions with recalling emotion from contextual information derived from the background of the picture. We believe this conflict becomes less prevalent as more time elapses since capture. We further used the selfie cue in additional studies as a successful metric to evaluate user satisfaction with mobile phone applications,[8] and we even proposed unobtrusively measuring drivers' user experience during their commutes.[9]

Re-Live the Moment

We investigated whether using visual as well as audible memory cues (pictures and music) captured during a beneficiary activity (such as running) could be used to facilitate the formation of positive habits (such as exercising often). In fact, contemporary psychology has shown that people are more likely to form a habit if they are reminded of previously positive experiences during habit formation. Based on this theory, we developed Re-Live the Moment (see Figure 2b), a mobile application that captures pictures and records the music that a user was listening to while on a run to create a personal exercise "music clip."[10] These clips act as video memory cues that can later be watched to remind runners of the positive feelings exhibited during their run, thus encouraging them to continue exercising.

We tested our prototype in a pilot study with five participants who reported that they enjoyed reviewing the multimedia presentation (the personal music clip) after the run, and some even shared the resulting clip with friends. We assume that positive feelings exhibited during the review might lead to more exercising, but a larger deployment, and for a longer duration, is required to ascertain the presence and significance of such an effect.

Déjà Vu

Human memory isn't restricted to simply recalling the past (episodic memory); it also relates to remembering events that are scheduled to occur in the future. In fact, human memory relies on prospective memory for remembering upcoming events.

To tap into the potential of prospective memory, we envisioned exploiting the concept of déjà vu (see Figure 4c) and displaying information about upcoming events and situations with the goal of making future situations appear somewhat familiar.[11] New situations naturally create a sense of excitement or anxiety. However, using peripheral displays in people's homes to present small information chunks that possibly have future relevance for a person can help lower potential anxiety caused by the uncertainty of the unknown. We investigated whether people can learn incidentally and without conscious effort about new environments and other people. By providing visual information, such systems create a sense of déjà vu at the point when people will be facing a new situation.

Assessing the Tools

In the final stages of the Recall project, we're undertaking a series of trials to assess the effectiveness of our memory augmentation tools.

We have identified three domains: domestic, workplace, and campus.

In a *domestic* setting, we use devices and displays to present memory cues in accordance to people's regular routines. We attach displays in the periphery of their homes to create stimulating environments and display personal content with the goal of supporting people's cognition and memory.

In a *campus* scenario, we focus on the scheduling and presentation of personalized media on public displays and other ambient displays across a university campus. By deriving memory cues from lecture material, we target content at specific individuals and groups.

Finally, the *work* scenario involves a series of augmented meeting rooms that give people access to captured moments, both from other meeting participants and from the installed infrastructure (such as cameras). Captured moments are augmented with topics inferred from an automated topic analysis system and played back to attendees on peripheral displays (such as on laptop and smartphone lock screens, or on tablets installed in offices that serve as picture frames). The goal is to help users better remember the meeting progression and outcome to prepare for the next meeting.

Challenges and Lessons Learned

There are still significant challenges that must be addressed before we can fully realize the potential of multimedia to augment human memory recall. Modern lifelogging devices, for example, can certainly capture a tremendous amount of information, yet this information is often irrelevant to the actual experience we strive to capture. Although a variety of approaches can be taken to separate important moments from mundane ones during capture, there is clearly much to be done.

Capturing Meaningful Data

Memory cues can be highly effective when evoking fine-grained details about an experience, or completely obsolete when they can't be placed into a context. Obtaining meaningful cues from a multimedia capture relies heavily on both the raw data quality and the legitimacy of the conclusions we draw from them. This is especially true for implicitly collected data, where inferential conclusions might be ambiguous.

The final quality of a multimedia capture thus can't be judged simply by the data itself—for example, the picture quality or its contents (though lower boundaries, such as dark or blurry images, do exist). Memories and their corresponding cues are highly personal. A supporting system therefore must learn the user's preferred types of cues and subjective relevance of a capture. For some, a blurry image of scribbled meeting notes might be enough to recall the meeting's content, while others might need to see the faces of those present to evoke a meaningful memory of the event. The use of physiological sensing might offer some insight into which cues hold the most potential for a user.

Dealing with Technology Constraints

There are constraints on what can be tracked, especially when it comes to physiological and also psychological or emotional aspects. Despite the progress made regarding tracking physical data, tracking mental activities is inherently difficult to do without additional hardware, such as eye trackers or Electroencephalography (EEG) devices—both of which are (still) highly obtrusive to use. Furthermore, certain mental states are difficult to infer reliably, such as attention levels, emotions, or stress.

A much more straightforward technical barrier is today's often low capture quality. Although storage will continue to expand, the sheer volume of what we can capture might nevertheless tax effective local processing, requiring extensive offline processing that might eventually become cost effective with technological advances.

Addressing Privacy Implications

Although the continuous capture of (potential) multimedia memory cues might be a boon to human memory, it might also represent the bane of an Orwellian nightmare come true. The strong social backlash that many wearers of Google Glass experienced[12] is a potent reminder of the potentially underlying incompatibilities between those who capture and those who are captured.

In previous work,[13] we enumerated the key privacy issues of memory augmentation technology—issues that span a wide range of areas, from data security (secure memory storage, ensuring the integrity of captured memories) to data management (sharing memories with others) to bystander privacy (controlling and communicating capturing in public). We recently started work on creating an architecture that both enables the seamless sharing of captured multimedia data within colocated groups (for example, an impromptu work meeting or a chat over coffee) and features tangible objects to easily communicate and control what gets recorded and who can access the data.

Ultimately, Recall aims to lay the scientific foundations for a new technology ecosystem that can transform how humans remember to measurably and significantly improve functional capabilities while maintaining individual control. Our work in Recall has only begun to scratch the surface of this exciting new application area. **MM**

Acknowledgment

Project Recall is funded by the European Union in FP7 under grant number 612933.

References

1. C. Gurrin, A.F. Smeaton, and A.R. Doherty, “Lifelogging: Personal Big Data,” *Foundations and Trends in Information Retrieval*, vol. 8, no. 1, 2014, pp. 1–125.
2. A. Baddeley et al., *Memory*, Psychology Press, 2009.
3. K. Wolf et al., “Effects of Camera Position and Media Type on Lifelogging Images,” *Proc. 14th Int’l Conf. Mobile and Ubiquitous Multimedia*, 2015, pp. 234–244.
4. E. Niforatos et al., “PulseCam: Biophysically Driven Life Logging,” *Proc. 17th Int’l Conf. Human-Computer Interaction with Mobile Devices and Services Adjunct*, 2015, pp. 1002–1009.
5. E. Niforatos, M. Langheinrich, and A. Bexheti, “My Good Old Kodak: Understanding the Impact of Having Only 24 Pictures to Take,” *Proc. 2014 ACM Int’l Joint Conf. Pervasive and Ubiquitous Computing: Adjunct Publication*, 2014, pp. 1355–1360.
6. H.V. Le et al., “Impact of Video Summary Viewing on Episodic Memory Recall: Design Guidelines for Video Summarizations,” *Proc. 34rd Ann. ACM Conf. Human Factors in Computing Systems*, 2016.
7. T. Dingler et al., “Reading-Based Screenshot Summaries for Supporting Awareness of Desktop Activities,” *Proc. 7th Augmented Human Int’l Conf.*, 2016, pp. 27:1–27:5.
8. E. Niforatos and E. Karapanos, “EmoSnaps: A Mobile Application for Emotion Recall from Facial Expressions,” *Personal and Ubiquitous Computing*, vol. 19, no. 2, 2015, pp. 425–444.
9. E. Niforatos et al., “eMotion: Retrospective In-Car User Experience Evaluation,” *Adjunct Proc. 7th Int’l Conf. Automotive User Interfaces and Interactive Vehicular Applications*, 2015, pp. 118–123.
10. A. Bexheti et al., “Re-Live the Moment: Visualizing Run Experiences to Motivate Future Exercises,” *Proc. 17th Int’l Conf. Human-Computer Interaction with Mobile Devices and Services Adjunct*, 2015, pp. 986–993.
11. A. Schmidt et al., “Déjà Vu—Technologies that Make New Situations Look Familiar: Position Paper,” *Proc. 2014 ACM Int’l Joint Conf. Pervasive and Ubiquitous Computing: Adjunct Publication*, 2014, pp. 1389–1396.
12. B. Bergstein, “The Meaning of the Google Glass Backlash,” *MIT Technology Rev.*, 14 Mar. 2013; www.technologyreview.com/s/512541/the-meaning-of-the-google-glass-backlash.
13. N. Davies et al., “Security and Privacy Implications of Pervasive Memory Augmentation,” *IEEE Pervasive Computing*, vol. 14, no. 1, 2015, pp. 44–53.

Tilman Dingler is a researcher at the University of Stuttgart, Germany. Contact him at tilman.dingler@vis.uni-stuttgart.de.

Passant El Agroudy is a researcher at the University of Stuttgart, Germany. Contact her at passant.el.agroudy@vis.uni-stuttgart.de.

Huy Viet Le is a researcher at the University of Stuttgart, Germany. Contact him at huy.le@vis.uni-stuttgart.de.

Albrecht Schmidt is a professor of human-computer interaction at the University of Stuttgart. Contact him at albrecht.schmidt@vis.uni-stuttgart.de.

Evangelos Niforatos is a researcher at Università della Svizzera Italiana (USI), Lugano, Switzerland. Contact him at evangelos.niforatos@usi.ch.

Agon Bexheti is a researcher at Università della Svizzera Italiana (USI), Lugano, Switzerland. Contact him at agon.bexheti@usi.ch.

Marc Langheinrich is an associate professor at the Università della Svizzera Italiana (USI), where he works on privacy and usability in pervasive computing systems. Contact him at langheinrich@ieee.org.