

Visual Lifelogging in the Era of Outstanding Digitization

Mariella Dimiccoli, Petia Radeva

Barcelona Perceptual Computing Lab (BPCL),
Computer Vision Center (CVC) and University of Barcelona (UB)
{mariella.dimiccoli,petia}@cvc.uab.edu

Abstract. In this paper, we give an overview on the emerging trend of the digitized self, focusing on visual lifelogging through wearable cameras. This is about continuously recording our life from a first-person view by wearing a camera that passively captures images. On one hand, visual lifelogging has opened the door to a large number of applications, including health. On the other, it has also boosted new challenges in the field of data analysis as well as new ethical concerns. While currently increasing efforts are being devoted to exploit lifelogging data for the improvement of personal well-being, we believe there are still many interesting applications to explore, ranging from tourism to the digitization of human behavior.

1 Introduction

We are already living in the world, where digitization affects our daily lives and socio-economic models thoroughly, from education and art to the industry. In essence, digitization is about implementing new ways to put together physical and digital resources for creating more competitive models. Recently, lifelogging appeared just as another powerful manifestation of this digitization process embraced by people at different extents. Lifelogging refers to the process of automatically, passively and digitally recording our own daily experience, hence, connecting digital resource and daily life for a variety of purposes. In the last century, there has been a small number of dedicated individuals, who actively tried to log their lives. Today, thanks to the advancements in sensing technology and the significant reduction of computer storage cost, one's personal daily life can be recorded efficiently, discretely and in hand-free fashion (see Fig. 1). The most common way of lifelogging, commonly called visual lifelogging, is through a wearable camera that captures images at a reduced frame-rate, ranging from 2 fpm of the Narrative Clip to 35 fps of the GoPro. The first commercially available wearable camera, called SenseCam, was presented by Microsoft in 2005 and during the last decade, it has been largely deployed in health research. As summarized in a collection of studies published in a special theme issue of the American Journal of Preventive Medicine [5], information collected by a wearable camera over long periods of time has large number of potential applications, both at individual and population level. At individual level, lifelogging can aid in contrast dementia by cognitive training based on digital memories or in improving well-being by monitoring lifestyle. At population level, lifelogging could be used as an objective tool for

understanding and tracking lifestyle behavior, hence enabling a better understanding of the causal relations between noncommunicable diseases and unhealthy trends and risky profiles (such as obesity, depression, etc.)



Fig. 1. Evolution of wearable camera technology. From left to right: Mann (1998), GoPro (2002), SenseCam (2005), Narrative Clip (2013).

However, the huge potential of these applications is currently strongly limited by technical challenges and ethical concerns. The large amount of data generated, the high variability of object appearance and the free motion of the camera, are some of the difficulties to be handled for mining information from and for managing lifelogging data. On the other hand, legality and social acceptance are the major ethical challenges to be faced. This paper discusses these issues and it is organized as follows: in the next section, we give an overview of potential applications; in section 3, we analyze technical challenges and current solutions. Section 4 is devoted to ethical issues and, finally, in section 5, we draw some conclusions.

2 Potential Applications

Humans have always been interested in recording their life experiences for future reference and for storytelling purposes. Therefore, a natural application would be summarizing lifelog collections into a story that will be shared with other people, most likely through a social network. Since the end-users may have very different tastes, storytelling algorithms should incorporate some knowledge of the social context surrounding the photos, such as who the user and the target audience are. However, lifelogging technology allows capturing our entire life, not only those moments that we would like to share with others (see Fig. 2). This offers a great potential to make people aware of their lifestyle, understood as a pattern of behavioral choices that an individual makes in a period of time. This feedback could provide education and motivation to improve health trends, detecting risky profiles, with a personal trainer “in-the-loop”. Indeed, by providing a symbiosis between health professionals and wearable technology, it could be possible to design and implement individualized strategies for changing behavior. Considering that physical activity and poor diet are major risk factors for heart diseases, obesity and leading causes of premature mortality, this social impact of applications will be huge. On the other hand, lifelogging could be useful in monitoring patients affected by neurological disorders such as depression or bipolar disorder by aiding in predicting crisis.



Fig. 2. Images recorded by a Narrative Clip: From left to right and from the 1st to the 2nd row: in a bus, biking, attending a seminar, having lunch, in a market, in a shop, in the street, working.

Finally, digital memories could be used as a tool for cognitive training for people affected by Mild Cognitive Impairment (MCI), a condition that represents a window for novel intervention tools against the Alzheimer disease. Although the emphasis nowadays is on the use of wearable cameras for health applications, its potential spreads to many other domains ranging from tourism to digitization of intangible heritage. For instance, data collected during a long trip could be used to make short and original photostreams for storytelling purposes and be shared in a network of visitors of a country. On the other hand, probably in the next century, these data would be useful for people interested in comparing how transportation and landscape have changed over time. During the last few decades, there has been an increasing interest in the use of digital media in the preservation, management, interpretation, and representation of cultural heritage. Intangible cultural heritage consists of non-physical aspects of a particular culture, among which folklore, traditions, behavior. The intangible aspects of our cultural heritage represent a treasure of significant historical and socio-economic importance. Naturally, intangible cultural heritage is more difficult to preserve than physical objects. The digital documentation of intangible cultural heritage represents a huge market potential, which is largely unexplored. Wearable cameras could be used in this field to collect, preserve and make available digitally part of the intangible cultural heritage of the 21th century, such as human behavior.

3 Technical Challenges

Wearing a camera over a long period of time generates a large amount of data (up to 70.000 images per month), making difficult the problem of retrieving specific information. Beside data organization, the high variability of object appearance in the real world and the free motion of the camera make state of the art object recognition algorithms to fail. In Fig. 3 are shown two sequences acquired by wearing a Narrative Clip (2fpm): one can appreciate the frequency of abrupt changes of the field of view even in temporally adjacent images that makes motion estimation unreliable and frequent occlusions that cause important drop in object recognition performances.



Fig. 3. Example of photostreams captured by a Narrative CLip while (first row) biking and having a coffee (second row).

As shown in [2], the interest of the computer vision community is rapidly increasing and this trend is expected to continue in the next years. Most available works have been conceived to analyze data captures by high temporal resolution wearable cameras, such as GoPro or Google Glasses and they can be broadly classified depending on the task, they try to solve in: activity-recognition [15, 11, 10, 13, 6], social interaction analysis [1, 3, 19], summarization [4, 16, 12]. Activity recognition usually relies on cues such ego-motion [15, 10], object-hand interaction [11, 10] or attention [13, 6]. Generally, the major difficult to be faced in the task of activity recognition are the large variability of objects and hands and the free motion of the camera that make it very difficult to estimate body movements and attention. Social interaction detection is based on the concept of F-formation that models orientation relationships of groups of people in space. F-formations require estimating pose and 3D-location of people, which are challenging tasks due the continuous changes of aspect ratio, scale and orientation. A common approach to summarization is to try to maximize the relevance of the selected images and minimize the redundancy. Relevancy can be captured by relying on mid-level or high-level features. Mid level features may be motion, global CNN features [4, 16], whereas high-level features may be important objects [12] or topics [18].

4 Ethical Issues

Lifelog technology can be considered still in its infancy and assuring that the related ethical issues receive full consideration at this moment is crucial for a responsible development of the field. In the last few years, a number of papers has tried to inquiry into the ethical aspects of lifelogs held by individuals [17, 7, 14], discussing issues to do with privacy, autonomy, and beneficence. Images captured by a wearable camera clearly impact the *privacy* of lifeloggers as well as of bystanders captured in such images. In [7], the authors identified various factors to make a photo sensitive and proposed to embed into the devices an algorithm that use these factors to automatically delete sensitive images. The most general meaning of *autonomy* is to be a law to oneself. The authors of [8] recognize that lifelogging offers a great opportunity towards autonomy, since it allows to better understand ourselves. Moreover, they provide recommendations and guidelines to meet the challenges that lifelogs poses towards autonomy. *Beneficence* concerns with the responsibility to do good by maximizing the benefits to an individual or to society, while minimizing harm to the individual. A critical component is informed consent that should be signed by participant to research projects or clinical projects. More general specifications for wearable camera research are provided in [9], proposing an ethical framework for health research.

5 Conclusions

This paper has reviewed some of the most important aspects of visual lifelogging, focusing on the technical and ethical challenges it arises, and on its potential applications. We believe that a responsible development of the field could be highly beneficial for the society. In order to become widely used technology, a large amount of effort should be invested in the development of efficient information retrieval systems, to allow fast and easy access to lifelogging content at a semantic level. Further advances in the field of deep learning will allow filling this semantic gap.

Acknowledgments

This work was partially founded by TIN2012-38187-C03-01 and SGR 1219. M. Dimiccoli is supported by a Beatriu de Pinos grant (Marie-Curie COFUND action).

References

1. S. Alletto, G. Serra, S. Calderara, F. Solera, and R. Cucchiara. From ego to nos- vision: Detecting social relationships in first-person views. Proceedings of CVPR, pages 594–599. IEEE, 2014
2. A. Betancourt, P. Morerio, C. S. Regazzoni, and M. Rauterberg. The evolution of first person vision methods: A survey. IEEE Transactions on Circuits and Systems for Video Technology. To appear 2015.
3. V. Bettadapura, I. Essa, and C. Pantofaru. Egocentric field-of-view localization using first-person point-of-view devices, Proceedings of WACV, IEEE (2015)

4. M. Bolaños, R. Mestre, E. Talavera, X. Giró-i Nieto, and P. Radeva. Visual summary of egocentric photostreams by representative keyframes. *Proceeding of WEsAX*, IEEE, 2015.
5. A. R. Doherty, S.E. Hodges, A.C King, A.F. Smeaton, E. Berry, C.J. Moulin, S. Lindley, P. Kelly, and C. Foster. Wearable cameras in health: the state of the art and future possibilities. *American journal of preventive medicine*, 44(3):320–323, March 2013.
6. A. Fathi, Y. Li, and J. Rehg. Learning to recognize daily actions using gaze. *Proceedings of ECCV 2012*, pages 314–327. Springer, 2012.
7. R. Hoyle, R. Templeman, S. Armes, D. Anthony, D. Crandall, and A. Kapadia. Privacy behaviors of lifeloggers using wearable cameras. *Proceedings of Ubicomp*, pages 571–582. ACM, 2014.
8. T. Jacquemard, P. Novitzky, F. OBrolch’ain, A. Smeaton, and B. Gordijn. Challenges and opportunities of lifelog technologies: a literature review and critical analysis. *Science and engineering ethics*, 20(2):379–409, 2014.
9. P. Kelly, S. J. Marshall, H. Badland, J. Kerr, M. Oliver, A. R Doherty, and C. Foster. An ethical framework for automated, wearable cameras in health behavior research. *American journal of preventive medicine*, 44(3):314–319, 2013.
10. S. Lee, S. Bambach, D. Crandall, J. Franchak, and C. Yu. This hand is my hand: A probabilistic approach to hand disambiguation in egocentric video. *Proceedings of CVPR*, pp 557–564. IEEE, 2014.
11. C. Li and K. Kitani. Pixel-level hand detection in ego-centric videos. *Proceedings of CVPR*, pages 3570–3577, IEEE, 2013.
12. Z. Lu and K. Grauman. Story-driven summarization for egocentric video. *Proceedings of CVPR* pages 2714–2721. IEEE, 2013.
13. K. Matsuo, K. Yamada, S. Ueno, and S. Naito. An attention-based activity recognition for egocentric video. *Proceedings of CVPRW*, pages 565–570. IEEE, 2014.
14. T.M. Mok, F. Cornish, and J. Tarr. Too much information: visual research ethics in the age of wearable cameras. *Integr. Psychological and Behavioral Science*, 49(2):309–322, 2014.
15. Y. Poleg, C. Arora, and S. Peleg. Temporal segmentation of egocentric videos. *Proceedings of CVPR*, pages 2537–2544. IEEE, 2014.
16. E. Talavera, M. Dimiccoli, M. Bolaños, M. Aghaei, and P. Radeva. R-clustering for egocentric video segmentation. In *Pattern Recognition and Image Analysis*, pages 327–336. Springer, 2015.
17. Klaus Wolf, Albrecht Schmidt, Agon Bexheti, and Marc Langheinrich. Lifelogging: You’re wearing a camera? *Pervasive Computing*, IEEE, 13(3):8–12, 2014.
18. M. Schinas, S.n Papadopoulos, Y. Kompatsiaris, and P. A. Mitkas. Visual event summarization on social media using topic modelling and graph-based ranking algorithms. *Proceedings of MIR*, ACM, 2015.
19. M. Aghaei, M.Dimiccoli, P. Radeva. Multi-face tracking by extended bag-of-tracklets in Egocentric videos. arXiv preprint. arXiv:1507.04576