An Examination of a Large Visual Lifelog

Cathal Gurrin, Alan F. Smeaton, Daragh Byrne, Neil O'Hare, Gareth J.F.Jones and Noel O'Connor

Adaptive Information Cluster, Dublin City University, Dublin 9, IRELAND
cgurrin@computing.dcu.ie

Abstract. Lifelogging is the act of recording some aspect of your life in digital format. A basic and common form of lifelogging is the creation and maintenance of blogs, which are typically textual in nature, though often with multimedia elements. In this paper we are concerned with visual lifelogging, a new form of lifelogging based on the passive capture of photos of a person's experiences. We examine the nature of visual lifelogs, and the differences between visual lifelog photos and explicitly captured digital photos. This is done by examining a million lifelog photos, or a year of visual lifelog data.

1 Introduction & Background to Visual LifeLogging

LifeLogging is the process of digitally capturing ones life experiences and the most popular form of this is to record a text description of some part of your day in a blog. Most blogging activities are text-only, though increasingly we are seeing bloggers include visual aspects such as deliberately taken digital photos or video clips. These are usually included to illustrate some aspect of the blog such as "*Here is a picture of the place I visited*" or "*Here is a picture of my friend John*". Such inclusion of deliberately posed and deliberately taken photos/videos is distinct from passively taken photos/images which constantly record the wearer's activities, visually. This is somewhat similar to a personal CCTV system, worn by the wearer, for the wearer's own, exclusive use. We call this visual lifelogging. Despite its relative novelty, visual lifelogging is gaining popularity thanks to projects like the Microsoft SenseCam [1] and Reality Mining [2]. Such lifelog data can include text, visual information (video or photos), audio information, biometric data (heart rate, galvanic skin response, blood pressure, etc.), location data, co-presence information and more besides.

In this paper we examine a collection of photos gathered by the constant wearing of a visual lifelogging device (SenseCam) for a period of more that one year. We explore the composition of these photos and compare how photos from visual lifelogs differ from conventional personal digital photo collections. The motivation for this work stems from fact that much research is ongoing into content analysis of digital photos, while in the absence of large-scale visual lifelogging efforts, there has been

little research undertaken for similar analysis of visual lifelogs. As during a typical day, the wearer of a SenseCam will passively capture up-to 3,000 photos, the need for automatic organization of these photo collections is compelling.

1.1 Visual Search & Retrieval of Photos

There has been much research recently on the organization of personal photo archives [3,4,5]. Some of this research aims to organize personal photo archives by exploiting the results of visual content analysis of the photo in order to extract some semantic meaning, with the goal of aiding the organization, search and retrieval of the photos. An example is facial analysis of photo content, to enumerate faces or to match faces across an individual's entire photo archive [6]. In addition, photo retrieval research has exploited the context of photo capture, or in some cases both visual content and photo context [4] to aid organisation. Often content analysis of photos requires the development of sets of concept detectors (e.g. face, crowd, building, etc…) which are trained on a representative set of photos [8]. In this paper we compare and contrast passively-captured large-scale visual lifelog data with more traditional intentionally-captured personal photo collections. We do this in order to explore the nature of a visual lifelog and to aid our understanding of the contents and composition of such a collection. Consequently we hope this understanding will allow us to leverage and deploy existing knowledge and techniques from the management and content analysis of personal photos in order to aid in automatic organization of visual lifelogs.

1.2 Visual LifeLogging

The device used in this research is a Microsoft SenseCam [7], which is a small wearable digital camera (worn around the neck) that is designed to take photos passively (without user intervention). Unlike a digital camera, the SenseCam has a fisheye lens, to maximize the field of view, and incorporates multiple sensors including; light sensors (intensity and colour), a multi-axis accelerometer, a thermometer and a passive infra red sensor to detect the presence of a person. Used collectively these sensors can trigger the capture of a photo. If capture is not triggered by one of these sensors, by default the SenseCam will take a photo after 30 seconds. This means that a SenseCam will normally capture approximately 3,000 images per day, amounting to one million images per year. For this research the author wore the device constantly over one year and amassed a large-scale visual lifelog containing over a million images. The device was worn constantly throughout the day, from breakfast until sleep. Sample photos are shown in Figure 1 to illustrate the nature of the visual lifelog data captured in this experiment. Typically in visual lifelogging research, participants wear a portable camera only for short periods of time, to record single activities or significant events. Most importantly and uniquely, this is the first time an individual has worn a SenseCam for such a prolonged period of time. The potential benefits of a personal visual diary such as that generated by a SenseCam have been detailed by Hodges et al. [5] and include the maintenance of personal

histories, security benefits by maintaining mobile visual logs and healthcare benefits, both for healthcare professionals and patients.

(a) Back View,   (b) vehicle scene,   (c) conversation scene, (d) screen scene
Figure 1. Typical Visual Lifelog Photos

2 Analysis of a Large Visual Lifelog

The process of manually analysing all one million photos in the visual lifelog collection would be too time consuming, therefore one thousand photos were chosen at random for individual examination. These photos were examined for the presence of a range of concepts and compared to a conventional digital photo collection of over ten thousand photos, captured using standard digital cameras. The analysis in this paper does not focus on particulars of the capture device such as image size or lens quality, rather emphasis has been placed on visual content which has been captured passively compared to that captured intentionally by conventional digital camera.

2. 1 General Observations on Visual Lifelog Data

An observation of visual lifelog data captured using a wearable device such as a SenseCam, clearly shows the differences between such data and a conventionally gathered photo collection. First, there is the prominence of the wearer's hands and arms (51% of examined images contained visible hands and/or arms, as in Fig 1b, c & d). Second, in most cases, there is often no clear salient object (Fig 1b & d) in the photo (people/objects are often in the periphery), whereas with a conventionally posed photo, there is normally an identifiable salient object typically focused in the center of the image. In addition, the author's office/work environment amounted to 16% of photos (easily identified by the presence of a computer screen, see Fig 1d) while steering wheel photos (see Fig 1b) amounted to 15% of the total. Finally, it is important to note that as the lifelogging device was worn around the neck of the owner, and as a result, the photos noticably appear to be captured from below eye level, often producing 'headless' shots of individuals and/or an unusual viewpoint. Since the user does not have to initiate capture we often see scenes where the user is actively using both hands and during conversations individuals are often captured mid-gesture adding to the conversational style and reinforcing the 'naturalness' of the photos. This contrasts heavily from conventional photo capture where photos are often staged and individuals and objects 'arranged' prior to capture.

2. 2 Photo Quality Analysis

The quality of a photo is important for both end-user viewing and for content analysis algorithms. This is no exception for the analysis of visual lifelogs. Our assumption is that a conventional digital photo, requiring an explicit user action to capture, will convey rich semantic meaning and will be of higher quality compared with the automatically captured photos from a wearable passive capture device which doesn't possess a flash or adjustable focus. Each of the randomly selected 1,000 photos were manually evaluated on a five-point scale which ranged from very low quality to excellent or 'photo album' quality. The results are summarized in Table 1.

Table 1. Visual quality of lifelog photos.

Very Low	Poor	Reasonable	Good	Photo Album
14%	25%	35%	22%	4%

Very low and poor quality photos (40%) are likely to be of little use for people reviewing the happenings of a day and therefore could be automatically removed from the collection prior to organizing the day, into, for example a sequence of distinct events [7] for presentation to the user. In addition, these photos are less likely to be of benefit to content analysis tools and their removal could save processing resources when analysing content. Typical very low and poor quality photos are likely to be dark or blurred[1] photos where the content of the photo is not obvious by casually examining the photo. Of the remaining 60% of photos, the good photos (22%) are ideal candidates for presentation to users or for content analysis (e.g. all photos in Figure 1) in that they are not blurred and visually excellent. The photo album quality photos are visually excellent, but also convey semantic meaning, just like a traditionally posed photo from a photo album and we estimate that almost seventy such photos would be captured in a typical day wearing a Sensecam[2]. Finally, the 35% of photos that are of reasonable quality will visually acceptable, and one would consider that they would be also be suitable for content analysis.

2.3 Comparison to Digital Photo Collections

In addition to understanding the quality of the visual lifelog photos, it is important to determine how similar lifelog photos are to personal photos. This is especially important if we plan to deploy existing semantic concept detection techniques to visual lifelogs. Table 2 presents findings from examining the 1,000 random lifelog photos and comparing these to 500 randomly chosen photos from the personal photo collection of the author (2,869 photos) and also 500 photos the photo collections of

---

[1] The SenseCam employs accelerometers sensors to trigger photo capture when the likelihood of blurring is minimised. Without such sensors, more blurred content would be captured.

[2] We have found that the notion of semantic meaning in lifelog photos is a very subjective concept and with inter-annotator reliability being poor.

over ten people in our research group (10,523 photos). As can be seen, both conventional, personal photo collections are equivalent (with a small amount of variation), as expected, however, the visual lifelog photos are very different in nature to the personal photo collections. There is only one similar concept between the lifelog and the conventional photo collections, and that concept was 'photos containing people'. All other concepts are vastly different and clearly show the differences between visual lifelogs and conventional photo collections. The typical concept detectors that could be applied to personal photo collections such as building detectors, cityscape and landscape detectors are less likely to be applied to visual lifelog data. Rather we are presented with a whole new set of concepts that are important for visual lifelogs, concepts such as computer screens, conversations, vehicle scenes and work scenarios are typical of common concepts in lifelog data, which if identified can be used to help organise a visual lifelog.

Table 2. Comparison to Conventional Digital Photo Collections (values in percentages).

	LifeLog (1 million)	Personal Photos (2,869)	Multi-User (10,523)
People	29.9	15.0	30.5
Buildings	3.5	43.3	35.0
Indoor	73.4	9.4	15.2
Outdoor	5.1	90.6	84.8
CityScape	2.1	26.2	22
Landscape	1.1	23.2	23.8
Computer Screen/TV	**26.7**	**< 1.0**	**0.0**
Conversation Scenario	**13.3**	**< 1.0**	**< 1.0**
In a Vehicle	**16.2**	**0.0**	**0.0**
Work Scenario	**24.6**	**< 1.0**	**0.0**
Back view of People	**6.8**	**1.0**	**2.0**

Rows in bold are concepts that we know to be very specific to visual lifelogs and as expected, are virtually non-existent in more traditional photo collections due to the fact that people would not often take conventional photos of work scenarios, driving scenarios or rear-views of people. Our findings suggest that the focus of semantic concept detectors for visual lifelog data will need to be tailored for the unique nature of visual life logs. In addition, the vast presence of almost redundant 'junk-content' such as computer screens, steering wheels, gives scope for filtering and summarizing the lifelogs automatically.

3 Conclusions and Future Work

We have illustrated that the photos contained within a large-scale visual lifelog collection are vastly different in visual content to conventional digital photos. It is important to note that the collection examined consists of continuously collected photos for a single user for a one year period, we have no reason to believe that this collection is not typical of what one would expect to find across all lifelogs and we

anticipate that the major distinctions between traditional and lifelog photo collections highlighted in this work will hold true. We are, however, currently working to collect large scale visual lifelogs from a range of users to confirm these initial findings.

We have found that visual lifelog photos often don't have salient objects, many are either low quality or redundant, and the types of scenes/objects captured differ greatly from conventional photo collections. These findings will affect the focus of our future work in developing specific concept detectors appropriate to lifelogs (such as driving, eating, talking or working) and the results of this experiment will help us in the migration of pre-existing concept detection techniques from the domain of digital photo collections to visual lifelogs. However, simply applying such concept detectors to visual lifelog data will not be sufficient to organize the enormous amounts of data involved, at Dublin City University we are researching new methods to help users organize these vast visual lifelog collections, to summarise and highlight, to filter, search and recommend from visual lifelogs, a content source that grows by up to 3,000 photos every day for a single user.

Acknowledgements

References

1. Gemmell, J., Williams, L., Wood, K., Lueder, R. & Bell, G. Passive Capture and Ensuing Issues for a Personal Lifetime Store. In : CARPE – 2004 , New York, USA October 2004.
2. Eagle, N. & Pentland, A., Reality Mining: Sensing Complex Social Systems, Personal and Ubiquitous Computing, Volume 10, Issue 4, 2006
3. Rodden, K & Wood, K. How do People Manage their Digital Photographs? In SIGCHI-2003, Ft. Lauderdale, Florida, USA, 5-10 April 2003.
4. O'Hare N, Lee H, Cooray S, Gurrin C, Jones G, Malobabic J, O'Connor N, Smeaton A.F & Uscilowski B. MediAssist: Using Content-Based Analysis and Context to Manage Personal Photo Collections. In: CIVR2006, Tempe, AZ, 13-15 July 2006.
5. Graham, A., Garcia-Molina, H., Paepcke, A., & Winograd, T. Time as essence for photo browsing through personal digital libraries. In JCDL 2002. Portland, Oregon, USA, 2002.
6. Cooray, S., & O'Connor, N. A Hybrid Technique for Face Detection in Color Images. In: AVSS - 2005, Como, Italy, 15-16 September 2005.
7. Hodges, S., Williams, L., Berry E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., & Wood, K. SenseCam : A Retrospective Memory Aid. In : UbiComp - 2006, CA, 2006
8. Naphade, M. and Smith, J. R. Learning Visual Models of Semantic Concepts, IEEE International Conference on Image Processing, Barcelona 2003