

Exploring Narrative Presentation for Large Multimodal Lifelog Collections through Card Sorting

Daragh Byrne^{1,2} and Gareth J.F. Jones¹

¹ Centre for Digital Video Processing, Dublin City University, Glasnevin, Dublin 9, Ireland

²CLARITY: Centre for Sensor Web Technologies
{daragh.byrne, gareth.jones}@computing.dcu.ie

Abstract. Using lifelogging tools, personal digital artifacts are collected continuously and passively throughout each day. The wealth of information such an archive contains on our life history provides novel opportunities for the creation of digital life narratives. However, the complexity, volume and multimodal nature of such collections create barriers to achieving this. Nine participants engaged in a card-sorting activity designed to explore practices of content reduction and presentation for narrative composition. We found the visual modalities to be most fluent in communicating experience with other modalities serving to support them and that the users employed the salient themes of the story to organise, arrange and facilitate filtering of the content.

Keywords: Lifelog, SenseCam, storytelling, digital narrative, media selection, narrative presentation, card-sorting, montage.

1 Introduction and Background

Lifelogging, or the capture of rich personal life histories through digital technologies, opens up wholly new opportunities for construction of highly personal biographical computation narratives [5,6]. Lifelogs may contain huge volumes of data from complimentary modalities including: visual diaries captured using the SenseCam; access to and the content of mobile and desktop documents; geo-location and people present using contextual sampling; explicitly captured photos or videos; and online social media, such as Twitter [5,10]. Given that a lifelog contains a wealth of content and context from a range of modalities, we must carefully consider how to select the most appropriate content to the story and how to integrate this multimodal content into a coherent, aesthetically pleasing presentation. This is particularly challenging. Tomai and Forbus [13] note narrative presentation ‘*serves to extract and highlight those events, outcomes and conclusions which the presenter and observer find “interesting.”*’ As Barthes observes, not all of these events will be of the same importance to the plot and will contain both *cardinal* units, which can be seen as the core moments in narrative on which the plot is hinged, and *catalysers* or the surrounding relevant moments which fill the space between them [4]. However, it should not just recount a sequence of events, moments or happenings from the author’s past, but should attempt to also communicate elements of the experience [2].

While some researchers have already explored narrative presentation in lifelog collections, these existing studies do not employ or offer computational models or strategies for narrative generation, and are been relatively small-scale or cursory [10,11]. For example, Harper et al. [11] explored the manual composition of stories from SenseCam footage but emphasis was placed on reporting their utility in reflection rather than narrative construction. Brooks offers one model for computational narrative [3], however, Appan et al. [2] demonstrate that this and other such approaches might not translate effectively to personal storytelling, finding narratives generated with such computational models to be “*disjoint and not very meaningful.*” As such within lifelogs, coherence is a paramount consideration for computational approaches. Rhetorical Structure Theory (RST) offers a framework for the explanation and analysis of coherence through the composition and relations between nuclear and satellite elements. While originally designed for text, RST has been adapted both to the analysis and summarisation of multimedia collections [12]. Additionally, Cheong et al. [7] offer a narrative framework for game logs which makes provision for coherency checking.

Christian and Young note that there has been little effort to verify that the computational approaches for narrative generation align with the expectations of audience and their mental models [8]. This motivates our work. In this paper we investigate how multimodal lifelog archives containing photo, text, and context data can be brought together to best enable storytelling by observing the practices and methods employed by users during the composition of personal narratives.

2 Experiment Outline

In order to gain an understanding as to how multimodal narration of a lifelog might be achieved, we asked 9 individuals in our research group (7 male, 2 female; age range 20-50) to engage in a card sorting activity [9] in which they arranged, selected and reduced physical representations of 'artifacts' from an extended lifelog collection to produce a multimodal story (see [6] for full details).

Prior to the experiment, participants selected up to five stories relating to an interesting activity from their SenseCam collections. Examples included: attending and presenting at conferences; personal holiday or site-seeing; and socialising with friends/colleagues. Participants were requested to provide their SenseCam images, and any related digital content created, reviewed or accessed e.g. web page(s) related to a conference they attended; digital photos captured; etc. On average 1687.5 images were provided to author each story, which were temporally skimmed to between 75 and 150 images. Only 6 stories had digital photo content (on average 37 photos per story.) Most stories had some digital document content available, mainly emails, this was in low volumes (average 6 emails per story.) The investigator printed and assembled this digital material into card format [9]. To fully simulate an extended lifelog, contextual, emotional and thematic information, were provided through an oral retelling of the experience by the participant. Participants were presented with the artifacts as a set of small lightweight cards. They were instructed to use them to create a multimodal representation of their chosen story by first reducing the content to the

50 items following which they iteratively reduced the content from 50 to 25, 25 to 10, 10 to 5 and 5 to one single item. Participants were also asked to 'think-aloud' [14], paying particular attention to their reasons for including or removing items, thereby providing insight into their decision-making and content reduction practices.

3 Results and Discussion

We now present the findings from the card-sorting activity. Prior work contains full details on collections employed [6].

General Observations: The composition of the story can be broadly defined by the types of artifact chosen and the aspects of that story that an artifact communicates. Table 1 illustrates this for a subset of the stories, while Tables 2 and 3 provide further breakdowns. As the author's intended meaning and the nature of the story will have a huge impact on its final representation, both the selected artifacts and the aspects of the story embodied by them vary widely. For example, user 5's story 1 pertains to a party and so the chosen artifacts predominantly embody people present, while user 8's story 1 is of a site-seeing trip, large reflecting those places visited. While the end representation may vary, we found that the strategies and methods users employed in assembling a multimodal story from lifelog content tended to be highly consistent.

Content Fluency: The range of available multimedia content required the participant to contemplate the affordances of each of the media modalities in communicating the story or its *fluency* and how this might impact on the end retelling. Tables 1 and 3 illustrate that users employed different modalities to communicate specific aspects of the story, suggesting that the modalities may have very different fluencies within the retelling. Almost 50% of a typical story is composed from SenseCam images (see Table 2.) making it the dominant modality. However, this is more likely to be a result of the volume available. In the 6 stories with digital photos, this photo media becomes the preferred modality. While preferred, users repeatedly expressed that it was desirable to include SenseCam content to support the photo-media, to provide a secondary viewpoint on that moment or to narrate something not otherwise captured by the available photos. As capture is often 'staged', the explicitly captured photos tend not to embody a particular *action*, however the SenseCam provides 'in the moment' capture of activities such as walking, presenting, eating, etc. Text-based or document content was less frequently employed and typically in an expository capacity. For example, user 1 employed Twitter messages succinctly summarising their site seeing; and users 3, 5, and 8 included an agenda, itinerary or programme.

Thematic Coherence & Clustering: The participants' initial action with the cards was to identify the salient themes and elements of their story. By reviewing these cards, the participant quickly identified which of the themes (and by consequence) aspects of the story are important to its plot. Once a user had identified the focal aspects of the story, they then cycled through the remaining content, iteratively associating relevant content to related theme (see [6], Fig 1a.) By doing so, they confirm that there is appropriate content to support representation of each aspect of the story,

establish sufficient coverage of the plot as a whole and ensure the story's coherence to its themes. A story consisted on average of 3.38 explicitly mentioned thematic or contextual elements (see Table 2), highlighting their importance to lifelog-based storytelling. The most central components were people present and actions undertaken, e.g. giving a presentation, etc. (see Table 3.) Only in a minority of cases was specific reference to time information made. While often mentioned during the oral retelling, emotional/mental reactions were not typically utilized. In one case, weather (it being cold) was also mentioned.

Core and Anecdotal Elements: In creating a story it was anticipated that the participants would emphasise the selection of artifacts core to communicating the overall activity, however, often elements which expressed peripheral or anecdotal encounters were preserved. These moments while not core to the plot, were noted by the users to be important to the overall experience. By preserving such elements, the narrative becomes more personal, exemplifying the unique experience of the individual. This aligns well with Barthes's *cardinal* and *catalysing* narrative units [4].

Content Reduction, Filtering and Removal: Users were presented with a large number of multimodal content (minimum 75 items) which they iteratively reduced. The processes by which the available content was reduced were largely consistent across all users. Users first conducted a quick review of the stack of cards, from which they identified the key themes and/or artifacts. After which, they sampled the available cards categorizing the content based on utility in retelling their desired story. Items that were relevant but redundant, due to repetition, were typically removed unless they provided an alternative or salient viewpoint on the activity e.g. a SenseCam image which might compliment an explicitly captured photo. Finally, poor quality or irrelevant content was immediately discarded.

Content retained from this review was clustered around the major themes of the story ensuring that they were appropriately covered by the available content. Content was then reduced to meet the upper-bound in one of two ways: by removing an entire cluster or by examining the artifact clusters and removing relevant but overlapping content e.g. visually similar photos, or items made redundant by another artifact. This content reduction process was consistently applied across the iterative reductions. Expository content – the artifacts used to explain or provide context to the story – was most likely to be removed as the upper-bound was lowered. Participants stated that these factors were implicit based on the presence of other content and could consequently be assumed or inferred. For example, explicit labeling of a person's presence would not be required if it was obvious from a photo that they were present.

Layout and Presentation: Two major arrangements of the content were observed during the card sorting activity. In a minority of cases, where the stories could be considered more 'habitual', content was arranged as a chronological account. This was seen in particular with user's 4, 8 and 9. Within the experiment typically participants employed a theme-oriented montage-style layout in which items were placed concentrically around the most important items (see [6], Fig 1b.) This has parallels with the work of Ames & Manguy [1] and Appan et al. [2], and as such is a sensible visual presentational to pursue within lifelog-based digital narratives.

4 Conclusions & Future work

While the richness of a lifelog opens new opportunities in the creation of personal digital narratives, its volume and multimodal nature pose significant challenges. We have explored narrative selection and presentation of lifelog content through a card-sorting activity. The observed practices were largely consistently across users. Participants identified the core themes from the content, removing low-salience or quality content, and clustering content around the central themes and reduced the content based on repetition and redundancy. The presentation took a montage or thematic layout. It is possible that observed consistencies could be due to the candidate population, and so we plan to conduct investigations to with a broader population to further validate the results. We also plan to conduct a more detailed quantitative analysis and explore the relations between chosen story elements using a framework such as RST [12]. It is anticipated that this will allow appropriate computational models for narrative generation within multimodal lifelog collections to be uncovered and implemented.

Acknowledgements: This work is supported by the Irish Research Council for Science Engineering and Technology, by Science Foundation Ireland under grant 07/CE/I1147. We would like to extend our thanks to the participants who took part in these experiments.

References

1. Ames, M., Manguy, L.: PhotoArcs: a tool for creating and sharing photo-narratives. In: Extended Abstracts CHI '06. pp. 466-471. ACM, New York, NY (2006)
2. Appan, P., Sundaram, H., Birchfield, D.: Communicating Everyday Experiences, In: Proc. SRMC 04, pp. 17-24. ACM, New York, NY (2004)
3. Brooks, K. Do Story Agents Use Rocking Chairs? The Theory and Implementation of One Model for Computational Narrative. In: Proc. Multimedia 1996, pp. 317-328. ACM (1997)
4. Barthes, R. Introduction to the structural analysis of narratives. Image, music, text; essays selected and translated by Stephen Heath. Hill and Wang, NY (1977).
5. Byrne, D., Jones, G.J.F.: Towards Computational Autobiographical Narratives through Human Digital Memories. In: Proc. SRMC 2008, pp. 9-12. ACM, New York (2008).
6. Byrne, D., Jones, G.J.F.: Creating Stories for Reflection from Multimodal Lifelog Content: An Initial Investigation. In: CHI 2009 Workshop on Designing for Reflection on Experience (2009)
7. Cheong, Y., Jhala, A., Bae, B.C., Young, M.R.: Automatically Generating Summary Visualizations from Game Logs. In: Proc. AIIDE 2008. AAAI, Menlo Park, CA (2008)
8. Christian, D., Young, R.M.: Comparing Cognitive and Computational Models of Narrative Structure. In Proc. AAAI 2004, pp. 385-390. AAAI Press, Menlo Park, CA (2004)
9. Fincher, S., Tenenberg, J.: Making sense of card sorting data. *Expert Systems*, vol. 22(3), pp. 89-93 (2005)
10. Gemmell, J., Aris, A., Lueder, R.: Telling Stories with MyLifeBits. In: Proc. of ICME 2005, pp. 1536-1539. IEEE Press, Washington D.C. (2005)
11. Harper R., Randall, D., Smyth N., Evans, C., Heledd, L., Moore, R.: Thanks for the Memory. In: Proc. BCS HCI 2007, vol. 2. pp. 39-42. BCS, Swinton, UK (2007).
12. Taboada, M. and Mann, W.C.: Applications of Rhetorical Structure Theory. *Discourse Studies*. vol. 8(4), pp. 567-588 (2006)

13. Tomai, E., Forbus, K.D. Narrative Presentation and Meaning, In: AAAI 2007 Fall Symposium on Intelligent Narrative Technologies, pp. 162-165. AAAI, Menlo Park (2007)
14. van Someren, M.W., Barnard, Y.F., Sandberg, J.A.C.: The Think Aloud Method: A Practical Guide to Modeling Cognitive Processes, Academic Press (1994)

Table 1. Example of lifelog artifacts selected by users to embody a chosen story within 5 items.

User	Story	Item 1	Item 2	Item 3	Item 4	Item 5
1	1	Twitter	SenseCam	SenseCam	SenseCam	IM Transcript
		<i>Exposition</i>	<i>Place</i>	<i>Place</i>	<i>Person</i>	<i>Exposition</i>
3	1	Photo	SenseCam	Photo	SenseCam	Document
		<i>Person</i>	<i>Person/Place</i>	<i>Person</i>	<i>Action</i>	<i>Exposition/Place</i>
4	2	Photo	SenseCam	SenseCam	Theme	Theme
		<i>People/Place/</i> <i>Object</i>	<i>People/Action</i>	<i>People</i>	<i>Object/Place</i>	<i>Object</i>
4	4	Photo	SenseCam	SenseCam	Photo	Theme
		<i>People/Object</i>	<i>People/Action</i>	<i>Action</i>	<i>Object</i>	<i>Time</i>
5	1	Photo	Photo	SenseCam	SenseCam	Photo
		<i>Person</i>	<i>Object</i>	<i>Action/Person</i>	<i>Action/Person</i>	<i>Person</i>
5	2	Photo	Photo	SenseCam	SenseCam	SenseCam
		<i>People/Place</i>	<i>Place</i>	<i>Action</i>	<i>Action</i>	<i>Action/Place</i>
7	2	Photo	Photo	Theme	Theme	Theme
		<i>Place/Object</i>	<i>Person</i>	<i>Emotion</i>	<i>Place/Object</i>	<i>Person</i>
8	1	Email	Sensecam	Sensecam	Sensecam	Sensecam
		<i>Exposition</i>	<i>Place</i>	<i>Place</i>	<i>Place</i>	<i>Place/Object</i>

Table 2. Summary of Content Types Employed (displays average number in the Top 10 media artifacts selected to embody a story).

Media Type	Totals	Stories with available photos	Stories with available text media
SenseCam	4.63	3.33	4.62
Photo	1.08	4.33	0.77
Theme	3.38	2.00	2.92
Document	0.08	0.17	0.15
Email	0.54	0.17	1.00
Web	0.17	0.00	0.31
Other	0.17	0.00	0.31
Number of Stories	24	6	13

Table 3. The aspects of a story embodied by the modalities a percentage of the total items (note that a single artifact may embody one or more aspect of the story, e.g. it might simultaneously represent a person and an action).

Aspect	All	Photo	Document	Context/Themes
Person	0.47	0.62	0.04	0.36
Place	0.19	0.38	0.04	0.19
Time	0.03	0.00	0.00	0.10
Action	0.32	0.08	0.04	0.06
Object	0.25	0.23	0.22	0.26
Emotion	0.03	0.00	0.00	0.08
Exposition	0.07	0.00	0.65	0.03
Other	0.00	0.00	0.00	0.01
Number of Items	240	26	23	78