Ambient Multi-Camera Personal Documentary

Richard Beales
IT Innovation Centre
2 Venture Road, Chilworth Science Park, Southampton, SO16 7NP, UK
+44 (0) 23 8076 0834
rmb@it-innovation.soton.ac.uk

Stuart E. Middleton
IT Innovation Centre
2 Venture Road, Chilworth Science Park, Southampton, SO16 7NP, UK
+44 (0) 23 8076 0834
sem@it-innovation.soton.ac.uk

Matthew Addis
IT Innovation Centre
2 Venture Road, Chilworth Science Park, Southampton, SO16 7NP, UK
+44 (0) 23 8076 0834
ma@it-innovation.soton.ac.uk

ABSTRACT
Polymnia is an automated solution for the creation of ambient multi-camera personal documentary films. This short paper introduces the system, emphasising the rule-based documentary generation engine that we have created to assemble an edited narrative from source footage. We describe how such automatically generated media can be integrated with and augment personally-authored images and videos as a contribution to an individual's personal digital memory.

1. INTRODUCTION
The explosions of interest in user-created digital content and life-caching have inspired a rash of tools aimed at simplifying the capture, annotation and subsequent retrieval and re-presentation of our personal experiences (e.g. [1][2]). Typically the digital photographs and home movies that these technologies handle are manually authored, by the individual themselves, or by friends or family members with whom they shared a particular experience, using handheld electronic devices. As such they present a familiar perspective - the same views that, a few years ago, might have been captured with a pocket 35mm camera, albeit without the processing costs and limited capacity of the analogue medium. Digital technologies also offer us the ability to move beyond the personal camcorder or cameraphone and document some significant events, such as family celebrations or visits to tourist attractions, from entirely new perspectives. Fitzgibbon and Reiter suggested one such scenario, where a birthday party in the home might be captured using ambient technologies and recreated as a 3D virtual environment [3]. In this short paper, we introduce the Polymnia system for ambient real-video documentation of visitors' activities within leisure destinations like amusement parks, art galleries and museums; particular emphasis is given to the documentary generation engine that we have developed to create an edited narrative from multi-camera source footage.

2. THE POLYMNIA SYSTEM
Polymnia [4] provides a complete solution for personal ambient video documentary creation. A chain of video cameras is positioned to provide well-framed views of visitors at key locations throughout the space. Live video from each camera is recorded continuously and deposited in a central AV repository. When a visitor registers with the system, photographic images of their face and full body are captured. As they then progress around the space, they are tracked as a moving region from one camera to another and their movements logged as industry-standard MPEG-7 metadata associated with the relevant video file. Where an individual does not pass directly from the field of view of one camera to that of another, or they are temporarily obscured within the frame, the face and body images captured at registration are used to re-associate their identity with the moving region. To add robustness to this reacquisition process, visitors can optionally also carry active RFID badges, the locations of which are used to maintain shortlists of visitors that might be expected to appear within the field of view of any particular camera. Live camera-switched feeds for each visitor registered with the system are available as secure web video streams that can be viewed remotely by friends or family.

As a visitor leaves the documented space, the automated editing process begins. All MPEG-7 metadata files associated with that individual are retrieved and compiled into a single MPEG-7 document that describes all recorded footage in which they appear. A set of pre-defined, venue-specific editing rules, created by a video professional with directorial expertise, is applied to achieve the most aesthetically pleasing edit possible with the available footage. Library footage, still frame inserts and graphic overlays, can also be incorporated according to the pre-defined rules. This editing process is described in more detail in section 3. To represent the resulting edit decision list (EDL) we have adopted the W3C's multimedia presentation language, SMIL 2.0[1]; the EDL can be played directly, using an appropriate viewer application with network access to the unedited source files, or rendered as a single movie file and written to a DVD as a standalone souvenir that the visitor can take away.

3. DOCUMENTARY GENERATION
Existing documentary generation systems structure known, pre-authored multimedia material to best fit a required narrative or presentation time (e.g. [5]).Our documentary generation engine was developed from a standpoint that it should support the unobtrusive documentation of natural behaviour within a space, not demand that individuals be shepherded from one camera to another to fit some prewritten script. Accordingly it needed to be flexible enough to accommodate both unexpected visitor behaviour, such as diversions from a suggested route, repeated activities or unusually fast or slow progress and technical difficulties such as an inoperable camera or temporary tracking failures.

3.1 A rule-based editing approach
The approach we have adopted draws inspiration from the fields of computational linguistics and expert systems. In syntactic parsing and sentence generation the grammatical structure of a

[1] http://www.w3.org/AudioVideo/

sentence is described by context-free rules that define valid word and phrase combinations according to their part-of-speech tags, e.g. verb, noun, article, verb phrase, noun phrase - a grammatically correct sentence is generated by applying the rules to a set of words to produce phrases which are aggregated to produce a complete sentence.

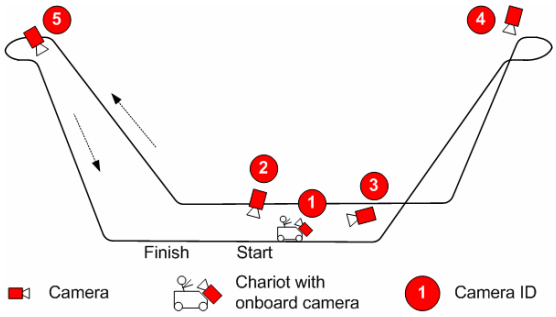

Figure 1 Roller coaster scenario camera layout

Figure 1 illustrates how this approach can be used to document an amusement park roller coaster ride. Cameras are positioned around the ride and onboard any chariots that will carry Polymnia users; Polymnia's vision-based tracking system detects when an individual appears within the field of view of any of the cameras. Figure 2 illustrates the shot sequences required to document this scenario. Three shots (S1, S2, S3) are combined to produce a film about a visitor embarking and starting the ride. Two short films of the climb and drop sequences are produced by concatenating shots S4, S5, S6 and shots S7, S8. These are combined to make a film about the ride. Finally, a film about the entire roller coaster experience is created by concatenating the short films covering embarkation, the ride itself, and disembarkation.

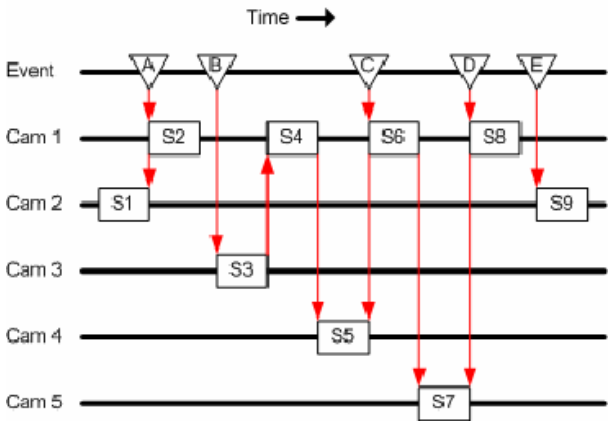

Figure 2 Roller coaster documentary structure

A rule-driven aggregation technique is used to compile the sentences, or video sequences, to construct a complete narrative. The roller coaster scenario illustrated in Figure 1 requires four aggregation rules. The first rule that defines an 'Embark' clip has three conditions describing the three atomic shots required to produce the clip based on the detection of a visitor in Cam 1. The second rule defines a 'Start' shot with start time equal to the end time of the 'Embark' shot and end time defined by the detection of the roller coaster chariot in Cam 3. The third rule defines the 'Drop' shot based on the end time of the 'Climb' shot and the detection of the roller coaster chariot dropping in Cam 4. The last rule concatenates the three shots to produce a film. As each rule is self-contained, individual aspects of a documentary template can be revised or rewritten without requiring wider restructuring of the template. The graphical user interface that has been developed to support a video professional in the development of these rules is illustrated in Figure 3.

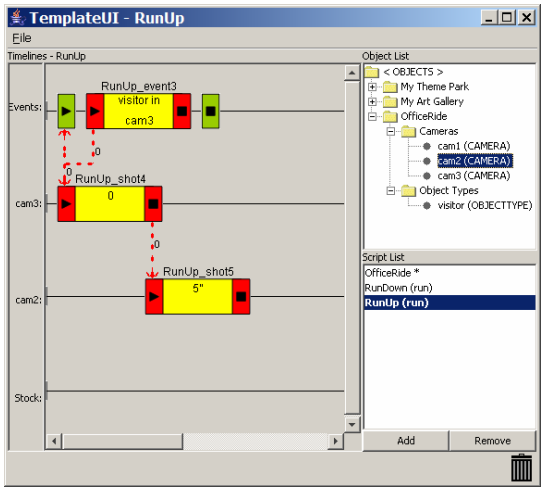

Figure 3 Editing rule creation tool

3.2 Annotation

One of the key challenges raised by the life-caching community is the need to adequately annotate user-created media so that select individual media items relevant to a particular narrative can be identified and retrieved without difficulty in the future. The emergence of ambient documentary technologies such as Polymnia will only add to the volume of digital material that individuals must manage, therefore it is essential that automatically generated content, too, is accompanied by accurate and comprehensive metadata. The SMIL Metainformation Module supports the inclusion of machine-readable annotation within our EDL, allowing RDF semantic descriptions to be associated with media elements or sequences of media elements. We exploit this feature not only to label the movie in its entirety with the location, time of visit and visitor identity, but also to identify key objects or activities (e.g. a roller coaster ride) within the edited sequence. In this way we support subsequent personal reuse of short sections of the movie, or even individual frames, in different narrative contexts. Like the video edits themselves, the application of metadata is specified by pre-authored rules.

3.3 Personal content integration

Just as a professionally created wedding video or a studio portrait of a child or family group is enjoyed alongside amateur snapshots and camcorder movies, so users of Polymnia-equipped venues will wish to audition with and incorporate into the automatically generated footage any photographs or videos that they themselves

captured during their visit. As a first step, we are developing the facility for individuals to upload to the system any digital photographs that they have taken during their visit, before the documentary creation process begins. The timestamps associated with these images will be used to place them within the generated narrative and display them alongside the ambient video in a split-screen presentation. Because the system is aware of the visitor's location at the moment they captured a particular image (provided they were being tracked on-camera at that time), we will also be able to provide metadata annotation for these personally authored images. An image captured while the visitor is being tracked on the roller coaster, for example, can be tagged with a description of that activity.

In the longer term, we envisage an individual's personal wearable devices will be coupled more tightly with the system for the duration of their visit. A cameraphone, for example, can be co-opted as a handheld video source or personal microphone; more ambitiously, we might seek to detect an individual's mood, so that only 'stimulating' or 'pleasing' experiences are included in the edit.

3.4 Evaluation

The documentary generation engine has been successfully tested as a standalone component using pre-recorded source video. The human identification and tracking modules have also been tested as standalone components. A live, full-system trial is to take place in the ALLOU Fun Park, Athens, Greece, during March 2007. We are exploring further opportunities for evaluation and practical deployment within the arts and cultural heritage sector.

4. CONCLUSION

The advances in digital imaging and storage technologies that have enabled users to manually document their lives in great detail also support the creation of ambient documentary spaces, where their activities can be captured automatically from multiple perspectives. Polymnia provides an automated solution for vision-based tracking and documentary creation, compatible with the life-caching community's need for personal media that is richly annotated and easily reusable.

5. ACKNOWLEDGEMENTS

POLYMNIA (IST-2-004357) is funded by the European Commission under FP6-2003-IST-2 strategic objective 2.3.2.7. Partners include ICCS/NTUA, IT Innovation, Telefonica I+D, Hewlett Packard, Joanneum Research, Giunti Interactive Labs, Cinegram S.A., Università Cattolica del Sacro Cuore and Hellenic Entertainment Park - ALLOU. Full details of the Polymnia project can be found at http://polymnia.pc.unicatt.it/.

REFERENCES

[1] Tuffield, M. et al., *Image Annotation with Photocopain*, in *Proceedings of First International Workshop on Semantic Web Annotations for Multimedia (SWAMM)*, Edinburgh, Scotland.

[2] Teixeira, D., Verhaegh, W., Ferreira, M., *An Integrated Framework for Supporting Photo Retrieval Activities in Home Environments*. In *Ambient Intelligence, LNCS 2875*, Springer, 2003.

[3] Fitzgibbon, A., Reiter, E., *Memories for life: Managing information over a human lifetime*. Grand Challenge proposal, published by UK Computing Research Committee (UKCRC), 2003.

[4] Doulamis, A., Kosmopoulos, D., Christogiannis, C., Varvarigou, D., *Polymnia: Personalised Leisure and Entertainment over Cross Media Intelligent Platforms*. In Hobson, P. et al., *Knowledge-Based Media Analysis for Self-Adaptive and Agile Multi-Media, Proceedings of the European Workshop for the Integration of Knowledge, Semantics and Digital Media Technology, EWIMT 2004, November 25-26, London, UK*. QMUL 2004.

[5] Rocchi, C., Zancanaro, M., *Adaptive Video Documentaries*, In *Proceedings of the HyperText Conference, August 26-30, 2003 Nottingham, UK*, pp. 36-37, 2003.