

## Indexing Audio-Visual Digital Media: the PathScape prototype

Mike Leggett

### Introduction

This paper reports some early stages of interdisciplinary research into relations between human memory and machine memory, investigating methods of storage and retrieval of media elements in the current context of information and communication technology (ICT). I propose an approach to indexing audio-visual media utilising a representational system that draws on a real-world time-space representation as the taxonomy for the indexing procedure. I describe and evaluate an interactive experimental prototype, *PathScape*, and outline further practice-based research approaches to author-defined storage and retrieval systems.

During John Sutton's introduction to the *Memory and Embodied Cognition Workshop* in November 2004 he made a passing and somewhat seasonal reference to "...being on the beach and having random thoughts cross one's mind...". The *PathScape* project which I describe here originates from that relaxed state. I was browsing a local history pamphlet that informed me the place I was enjoying was the first beach Captain Cook noted in his Journals as having indigenous people visible on the shore. The project developed as an approach to describing the many ways in which that part of the South Coast of NSW could be represented, and thus remembered.

The first prototype was put together by a small team which was, to use Andy Clark's term, "...constructing the niche..." (Clark, 2004) - we were making a model or mock-up that best visualised at the time the ineffable. We were using a recent technology to investigate a different mode of literacy and its relation to memory, utilising a representational system that drew upon a cinematic time-space representation as the taxonomy for the indexing procedure.

The approaches taken by the *PathScape* project will be described in the context of recent research in the 'memory machines industry' that seeks to develop tools for storing and retrieving audio-visual digital media, whilst accommodating the perceived needs of the 'memory worker', whether as an individual, or part of a closed or open group.

### Desperately seeking ... precedents and context

The Greek orators and rhetoricians, before the alphabet had been handed down, developed an elaborate form of artificial memory, described so fully in Yates' *Art of Memory*: "...a series of *loci* or places. The commonest, though not the only type of mnemonic place system was the architectural type .... We have to think of the ancient orator as moving in imagination through his memory building whilst he is making his speech, drawing from the memorised places the images he has placed on them" (Yates, 1966). Thus it could be claimed the first movies were a conceptual model made by the Greek rhetoricians. As he moved in his imagination through the *loci* contained by the imaginary building, in the mind's eye they were encountered as wide shots, tracking shots, panning, tilts, close-ups and flashbacks. Like the visual language of cinema developed in the 20th Century, this could be regarded as the very first 'classic' film narrative.

The notion of 'memory traces' and representations for and of recall, while remaining contested ground, forms the basis of memory storage and retrieval devices, from the dictionary to the encyclopedia, from the diary to the snapshot. Autobiographical and personal memory can be prompted by what Tulving terms "synergistic ephory" (Tulving, 1983) whereby the emotion or the memory is evoked or revived by means of a stimulus such as a sound, or a photograph. Often aided by the context of the recall, a writer for instance, through placement of artefacts or words in spatial relationship can create the circumstances which connect with the narrative (of a memory trace, event, object etc). We are not unfamiliar with the use of postcards and palm cards or scraps of paper placed around the room as a way of organising complex sources in the process of synthesising thoughts and events into fresh formulations. (The author witnessed recently two professional script-writers' working method, which involved them laying out palm cards and images around a studio, whilst working with a computer in the centre of the room to synthesis their content. Russell Crowe's portrayal of the schizophrenic John Nash in the movie *A Beautiful Mind* (2001) provides an image of this process in its pathological state.)

Central to academic pursuit and increasingly, the education and edutainment of the population, the ability to interact with external memory machines such as collections and libraries of knowledge, or evidence, located on computer servers around the globe is essential. Knowing where to go and how to search relies on a meshing of machine and human memory faculties.

### Indexing Options

Signs represent the present in its absence; they take the place of the present ... when the present does not present itself, then we signify, we go through the detour of signs. (Derrida, 1973)

Within the repositories of collected memory, in large public collections for instance, the stimulus relies on a common rather than private language of signs, most often expressed in a word index form. The machine-based memory industries which specialise in servicing this demand, by storing data and knowing how to retrieve it again, are continuing to develop the technologies with which to do it. Some are moving away from notions of information retrieval and database management towards information gathering, seeking, filtering and visualization (Schneiderman, 1998).

Indexing is a way to increase retrieval precision and accuracy by consistent application of subject terms in their preferred forms. ... A taxonomy is a controlled vocabulary presented in an outline view, also called a classified view or hierarchy. Terms are organized in categories reflecting general concepts (Top Terms), major groups (Broader Terms), and more specific concepts (Narrower Terms). The final terms at the end of a branch, often called nodes, can represent any specific instance of a Broader Term, including terms from an authority file of people, organizations, places, or things. (DataHarmony, 2000)

A taxonomy of indexing enables an overview of the topography of the system, by reducing scale and quantity to proportions that can be comprehended, particularly by new or inexperienced users. In many ways ideal for text-based data such as large ICT parallel database systems (Taniar & Rahayu, 2002), to approach audio-visual data based upon word interpretation is constraining, useful only when words and concepts in written documents need to be illustrated. Picture libraries use such an approach suggesting: "Search using keywords, concepts, image numbers, etc" (GettyImages, 2005). These systems are largely classified by human operators and interpretation is geared for the publishing and media industries.

Machine vision system development such as at the AT&T Laboratories in Cambridge have used image segmentation and neural net classifiers (Town & Sinclair, c.1998) to describe frame content ("classes of stuff") searchable by text or visual query. French Telecom commissioned work on Visual Information Retrieval (VIR) that used a similar semantic analysis based on the measurement of colour, texture and shape (Obeid, Jedynek, & Daoudi 2001). These machine methods are useful for the rapid and broad classification of the kind of images produced for surveillance or medical purposes, but inappropriate for specific collections related to specific themes or subjects with which the visitor interacts for the purposes of knowledge management.

The domain is of some interest to ICT manufacturers who develop text / keyword-based multimedia file management applications (such as *Extensis* and *Canto*) for aided retrieval. Hewlett Packard Labs in Palo Alto developed a prototype application for non-expert users – *Fotofile* – that "...blends human and automatic annotation methods." The approach assumed that the 'intuitive interface' would be a text-based annotation system at worst and a thumbnail browsing system at best. In the final outcome it combined the two together with the addition of some 'automatic' (machine) features. A crude face recognition feature which offered users matched faces to confirm and name, used a 'hyperbolic tree' diagram visual device to link each face with its occurrence in other images. The paper provided no quantitative assessment though some compelling qualitative comments:

Photography and home movies are activities that address deep human needs; the need for creative expression; the need to preserve memories, the need to build personal relationships with others. Digital photography and digital video can provide powerful and novel, ways for people to express, preserve and connect. However, the new technologies often raise new problems; the problems of multimedia organisation and retrieval... (Kuchinsky et al, 1999)

Browsing, as the means by which users match an image to memory or a perceived need, has itself been aided by the work of Lim, Smith and Lu from Monash University, who "...designed *i-Map*, an interactive system for visualising and navigating a large scale image database..." that by clustering images onscreen enabled the user to "...explore areas which look more promising..." before selecting an initial image which the system would then seek matches for before re-clustering (Lim, Smith, & Lu, 2004).

Relational models of this kind were described by Ballard and Brown in the early 1980s as turning away from *representing* models, to *matching* models from within a knowledge base. Thus proposition and inference became important aspects of interaction with the database (Ballard and Brown, 1982). A decade later Ballard used the term "personalised representations" (Ballard, 1991) to describe the means we use to facilitate everyday behaviour. Correctly identifying our toothbrush in a bathroom shared by the household is an example I suggest: some residents may use colour differentiation whilst others, distrustful of their colour memory, prefer placing their toothbrush in a part of the bathroom different to the others. Clark has described action-orientated representations "...that simultaneously describe aspects of the world and prescribe possible actions, and are poised between pure control structures and passive representations of external reality" (Clark, 1997).

The relational terms "more", "same", "less" are of interest in this context. These same words were used by child development researchers (Griffiths, Shantz, & Sigel, c.1968) continuing to develop Jean Piaget's Conservation Task, first defined in the 1940s. The OED defines 'conservation' in this context as a branch of psychology:

...faculty of conservation: memory proper, or the power of *retaining* knowledge, as distinguished from reproduction or reminiscence, the power of *recalling* it. (My emphasis)

'Memory industry' research projects so far outlined, have been primarily concerned with the rapid and 'automatic' storage of visual media using text for classification and thus retrieval purposes. Elsewhere, often with less formal research approaches, others have been exploring storage and retrieval of collections that move the representation of the storage system away from the textual toward utilising a time-space representational system, (though often with text-based augmentation). I offer a brief list:

- *Jerome B. Wiesner, 1915-1994: A Random Walk through the 20th Century*, from the Media Lab at MIT (Davenport & al, 1994). By monitoring the user's initial selection, subsequent options are reorganised to cluster related topics, using a combination of image and words and re-shuffling their relative positioning on the screen.

Web-based Thesaurus sites function in a similar way. (Thesaurus, 2004)

- The visual artist Chris Hales made *Twelve of My Favourite Things*, an interactive diaristic installation, accessed using a touch screen over an image composite of three Quicktime movies. Through interaction with 'hot spots' based on visible colour zones, movies narrating the world of some young children are accessed. (Hales, n.d.)
- *ArcView* is related to topography, recording time and place, and is used widely in industries related to environmental planning, water and land management, urban layout, national parks, mining and agriculture. These are specialised tool sets based on data derived from the scientific method of measurement using GIS satellite data combined with multimedia files. The City of Fairfield for instance, commissioned a system based on this tool, for relating the area's history, [Peopling Fairfield](#).
- the *Digital Songlines* project at the Australasian Centre for Interactive design (ACID) in Queensland similarly uses the graphical representations familiar in game engines to map the GIS data relevant to 'country', relevant to an indigenous community and its traditional and contemporary cultural activity in the Bush west of Brisbane. (Leavy, 2004)
- *Aspen Walk*, an MIT project from the late 1970s, linked two video disc players with a computer program to enable a 'walk' around the Rocky Mountains town, represented on a screen in front of the user, whereby the user/navigator could determine at each image of crossroads, to turn to the left, to right, or to carry straight on (Naimark, 1998.)
- *Swarm*, by installation artist Alex Davies, uses a series of video projectors, camera, and computer to create a data space that is constantly provisional, always in flux, your presence now absent, a previous presence now present. The space becomes charged with time as images of yourself and previous visitors are randomly stored and retrieved to be projected into the space of the gallery installation (Davies, 2003)
- The Exeter Cathedral website: "There are two main routes into the material, Visual and Verbal. .... The Verbal route is for those who are more at ease with text than images". There is an elegance and appropriateness in the visual part of the site that associates a contemporary on-line database visualization design with a medieval equivalent - the vaulting and keystones in a 700 year old cathedral, the pathways and nodes storing 15th Century arcane and local knowledge using visual coding and systematic method (Henry and Hulbert, 1998).

These systems use machine memory to store evidence for later retrieval, as a means of reasserting a past presence, the genesis of the *PathScape* project.

### PathScape

The *PathScape* prototype developed in 1998/2000 has an interface and navigation system giving access to 'narratives' by their association with a specific place or location or series of locations, like "the walk to work" or "the cliff-top walk". The taxonomy is represented with images of contiguous cinematic space - individual photo images are pixilated to produce apparent motion in a forward direction, perceived as a movement 'into' the space recorded, a landscape. The movement is achieved by gesture, using a mouse in the prototype, to control the on-screen cursor – see Figure 1.

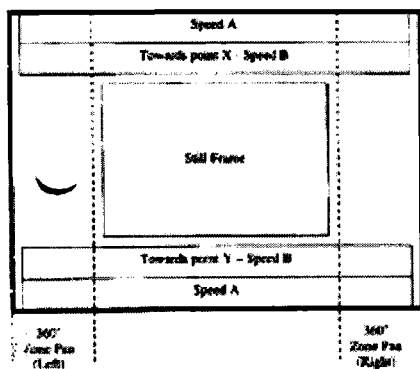
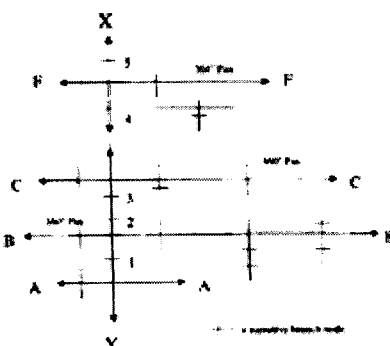


Figure 1: Screen Cursor Areas and Gesture Outcomes

The taxonomy of the Path is ordered by three indexical devices. The first is located in the border area that surrounds the central image of movement along the Path. Within this border are seen at various points, *fragments of images*, visible for short durations. These indicate a nodal junction which, when 'captured' by halting all apparent forward movement, enable with a click the launch of a movie to replace the image and sound of the Path. Thus along an X-Y axis are the 1, 2, 3, .... 8, 9 etc options, or loci 'in' which are stored the 'narratives' – see Figure 2.



-Figure 2; Schematic for accessing image/sound database.

The second device is changes in background *colour* to the border and background *sound*, signifying changes of zones. (Differences in ecology along the Path in this prototype). In Figure 2, along the X-Y axis are the AA, BB, CC .... FF etc axes. By gesturing to the left of the screen (or to the right) will launch a 360° panning movement, a movie representation of the zone through which the user is currently 'passing' - to the right will pan right, to the left will pan left. Within the pan will be 'found' further nodes to launch movies storing more narratives.

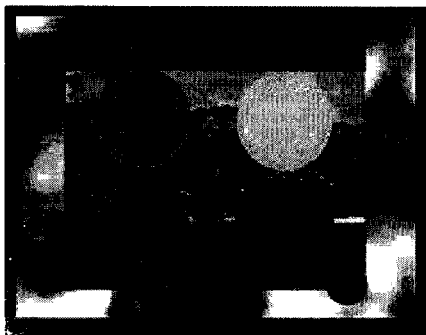


Figure 3: Screen grab within a narrative branch, with colour-coded circles.

At the completion of a narrative, the third indexical device appears as a series of *circle shapes* that appear over the final frame of the movie (Figure 3). Blue, yellow and brown and green circles function as 'buttons' to linked topics colour coded to symbolically represent a broad sort under the descriptors (in this prototype): Anecdotes, Historical Context, Commentary and Analysis. Each option extends and develops the background of what has gone before, in effect narrowing the index path to the specific, reducing from the broad.

Figure 4. *PathScape* system prototype demonstration – observe the position of the cursor to follow the navigation options being taken by and described by the author.

#### **Audience Feedback: specificity and further development**

The grey/black circles on the screen that sit behind each of the coloured circles are the route through to the traditional text-based index – the text information sits in the shadow, as it were, of its symbolic iteration as a movie. The text is organised sequentially as a series of 'browser pages' gathered, utilising XML protocols, from the Sources database of content, specifying

- for each narrative: Sound; Picture; Transcript; Keywords; Web Search;
- for the whole prototype: More Stories (as a Table of Contents - the narratives - with the frame numbers of the Path movie listed against each item, from which the narratives could be launched); Keyword Index

The user in the prototype therefore has a choice - to navigate the index by using images and sounds, or by using words, or a mixture of both.

The encounter in this prototype thus enables the user to orientate within a given topography in a way not dissimilar to a walk in the country or the city and to interrogate the surroundings for hidden evidence, for concealed information or comment, delivered as stories, as samples of discrete evidence of an individual or community.

### Planned Research

*Pathscape* is a project progressing through several stages and adopting several iterative forms. It can be delivered on disc (CD or DVD) or via the internet or broadband cable or conceivably, as it uses XML protocols, via a PDA or mobile phone. The system is dynamic, rebuilding the database at each launch, thus it is ideal for amending through the addition or removal of 'content' material.

With the further research into the development of appropriate interfaces that help the author(s) define the ontology and epistemology of personal and collective memory, the *PathScape* project will investigate a meta-design approach to placing and retrieving audio-visual digital media artefacts.

The aim is for users to engage through informed participation rather than being restricted to the use of existing systems. In such an event, this representational system will be open to invention by its author(s) through the placement of appropriate media into the chosen taxonomic indexing system. Different modes of taxonomic representation could be suggested in such a scenario to provide ways of thinking about the representation of memory.

The author(s) may not necessarily be wishing this to be for the benefit of what could be termed an audience, (anymore than diaries are for a wider audience), so there are many viewpoints from where evidence and its remembering could be approached: the author; the audience; the system function and design; or from the perspective of developing further links within interdisciplinary research.

The act of authoring – this could be termed *constructive action* – requires the relationship between the 'memory material' and its storage and retrieval to be addressed. Is the taxonomic and indexing approach to be achieved using visual media alone or in combination with words? In design building, how is the objective assessment of media collections (movies, stills, sound, graphics and the adaptability of 'newsreels') for drag-and-drop into the system to be determined? Or might the specific media composition of memory material for insertion into the system be a more effective solution?

When the author(s) are intending the system to be used by an audience beyond the authors or their group, then the following issues as examples, need to be considered from the audience's point of view – this could be termed *interpretive action*: taxonomy and indexing – visual and word semantics; presence – spatial orientation and navigation within the represented space and the spaces within which the system is encountered; the means by which gesture and control affect interaction with the system; the possibility for contribution to, feedback, or comment about the system, whereby the distinction between author and audience becomes blurred.

The system will have a database function and/or be a distributed system – this 'engine' would be largely hidden, but configurable as a generic, proprietary, prototype for distribution via disc or network.

The system could be 'tuned' for interdisciplinary research and encompass for instance, other mnemonic electronic devices such as PDAs, mobile phones, etc, or include sensor devices. Indeed content of the system could focus on specific topics such as the theory and science of memory.

### Conclusion

We can anticipate many more images being digitally authored and then consigned to the bottoms of drawers, for want of a means of retrieving their autobiographical or historical significance. In the contemporary context of tools like the Macintosh lifestyle tool *iLife*, audio-visual records could develop in usefulness to society, and with greater meaning to authors, if their presentation was orientated away from the linear forms of contemporary popular media which is dominant, toward the relational forms that have been the practice-based domain of many contemporary visual artists and increasingly, other disciplines too, which the PathScape prototype begins to demonstrate.

As the philosopher John Campbell has observed:

One of the most basic principles of plot construction is that the remembered 'I' traces a continuous spatio-temporal route through all the narratives of memory, a route continuous with the present and future location of the remembering subject. ... This principle imposes a kind of unity on all the narratives. (Campbell, 1997)

As collective or personal memory decays, the connectedness of events to the media artefact fade and the narrative thread is disrupted. The research project will investigate, using practice-based research (Leggett, 2005), means by which a system of indexing based on representations of place can assist users in locating media representations of memory or evidence of history. Meta-design will be explored as a means of enabling users to identify an indexical system appropriate to their placing of media elements that represent the present, past or conceivably, the future.

### References

- Ballard, D. (1991). Animate vision. *Artificial Intelligence*, 48, 57-86.
- Ballard, D., and Brown, C. (1982). *Computer vision*. New Jersey: Prentice-Hall.
- Campbell, J. (1997). The structure of time in autobiographical memory. *European Journal of Philosophy*, 5(5).
- Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, Mass: MIT Press.
- Clark, A. (2004). Beyond the flesh: enacting the boundaries while constructing the niche. Paper presented at workshops on memory and embodied cognition, Macquarie University, Sydney, November.
- DataHarmony (2000). Retrieved 1.9.04, from <http://www.dataharmony.com/faq.htm#b1>
- Davenport, G., & al. (1994). Jerome B. Wiesner, 1915-1994: a random walk through the 20th century. Retrieved 1.9.04, from <http://ic.media.mit.edu/projects/JBW/>
- Davies, A. (2003). Swarm. Retrieved 1.7.04, from <http://schizophonia.com/frmindex.htm>
- Derrida, J. (1973). *Differance - speech and phenomena*. NorthWestern University Press.
- GettyImages (2005). Gettyimages. 2005, from <http://creative.gettyimages.com/source/home/home.aspx>
- Griffiths, J., Shantz, C., & Sigel, I. (c.1968). *A methodological problem in conservation studies: The use of relational terms*. Lafayette: Merrill-Palmer Institute.
- Hals, C. (n.d.) Portfolio. Retrieved 1.9.04, from <http://www.smartlabcentre.com/4people/coreres/chaes.htm>
- Henry, A., & Hulbert, A. (1998). Exeter cathedral keystones and carvings. Retrieved 1.9.04, from <http://hds.essex.ac.uk/exetercath/index.html>
- Kuchinsky, A., Pering, C., Creech, M. L., Freeze, D., Serra, B., & Gwizdka, J. (1999). *Fotofile: A consumer multimedia organization and retrieval system*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit, Pittsburgh, Pennsylvania, United States.
- Leavy, B. (2004). *Digital songlines*. Brisbane: Australasian Centre for Interactive Design, QUT.
- Leggett, M. (2005). *Losers and finders: indexing audio-visual digital media*. Paper presented at the Creativity & Cognition conference 2005, Goldsmiths College London.
- Lim, S., Smith, R., & Lu, G. (2004). *I-map: An interactive visualisation and navigation system of an image database for finding a sample image to initiate a visual query*. Melbourne: Monash University.
- Naimark, M. (1998). *Place runs deep: Virtuality, place and indigenouness*. Paper presented at the Virtual Museums Symposium, Salzburg, Austria.
- O'Malley, M., Jedynek, B., & Daoudi, M. (2001). *Image indexing and retrieval using intermediate features*. Paper presented at the Proceedings of the ninth ACM international conference on Multimedia, Ottawa, Canada.
- OED. (2004). Oxford English Dictionary. Retrieved 1.9.04, from <http://dictionary.oed.com/>
- Schneiderman, B. (1998). *Designing the user interface*: Addison-Wesley.
- Taniar, D., & Rahayu, W. A. (2002). *Taxonomy of indexing schemes for parallel database systems*. Paper presented at the Distributed and Parallel Databases.
- Thesaurus. (2004). Visual thesaurus. 2004, from <http://www.visualthesaurus.com/online/index.jsp>
- Town, C., & Sinclair, D. (c.1998). *Content based image retrieval using semantic visual categories*. Cambridge: AT&T Labs Cambridge.
- Tulving, E. (1983). *Elements of episodic memory*. Oxford University Press.
- Yates, F. A. (1966). *The art of memory*. (1992 ed) Pimlico, London.