Visual Anthropology and Properties of the Medium (or The Visual Anthropologist in the Digital Library: From filmstrips to salient stills and back to Barthes).

David Zeitlyn

Department of Anthropology

Eliot College

The University

Canterbury, Kent CT2 7NS

Telephone +44 1227 764 000 #3360 Fax (44) 1227 827 289

Email: D.Zeitlyn@kent.ac.uk URL: http://lucy.kent.ac.uk/dz/

Abstract

Advances in digital image processing lead to ways of representing images and image-collections which approach the linguistic properties of concision and summary. However, important logical differences remain. Indexing is a form of summary representation with problematic application to visual material. 'Salient stills' and 'film strips' as representations of video clips have been suggested as solutions. Yet problems with abstract non-verbal representations return us to the theoretical stance of Walter Benjamin and Roland Barthes: images are approached textually.

Visual Anthropology and Properties of the Medium (or The Visual Anthropologist in the Digital Library: From filmstrips to salient stills and back to Barthes).

Introduction

If a picture is worth a thousand words what is the value of a digital video clip?

There are few generalizations as plainly wrong as proverbs, and few as patently incorrect as the suggestion that 'a picture is worth a thousand words' (which is not to deny the power of images, more to suggest that such power cannot be reduced to words). Recent changes in digital image processing have produced methods of summarizing which begin to approach those of conventional verbal description. It remains for visual anthropologists to make use of them. One consequence of the types of non-verbal summary being developed is that visual representations do *not* share many important features with language, concision and generalization being chief among them. This poses problems relating to the management of image collections, for both librarians and for users of the collections as well as for theoreticians struggling to find an appropriate analytic vocabulary. It provides a practical demonstration of what Sol Worth described so trenchantly in 1981: 'Pictures Can't Say "Ain't"'. For Worth, since pictures cannot express negative propositions they are not capable of truth or falsity, rather 'pictures are a way we structure the world around us. They are not a picture of it' (Worth 1981:182), or as MacDougall (1998:75) puts it 'while it is possible for a written ethnography to present an overall picture of cultural patterns in a community, a film can only present specific instances and imply some generality for them'. There is no way for a picture to express the propositional

statement ‘unicorns do not exist’. This is an interesting example since in the Fifteenth and Sixteenth Centuries Europeans were not sure whether or not this was the case. Indeed, early representations of the rhinoceros were influenced by an artistic tradition of the depiction of unicorns (as discussed by Gombrich 1960:70-1 see also Zeitlyn 1996). But the subtlety of Gombrich’s discussion of the role of convention in art does not affect Worth’s point: negative statements cannot be depicted. As Peacocke (1987:399) makes clear, there is no problem depicting non-existing entities such as unicorns or Don Quixote. The problem is representing their non-existence. Roskill and Carrier discuss this in greater depth in their monograph. However, from the start (1983:viii) they concede Worth’s basic point. They discuss Worth’s argument (1983:124 n. 23), suggesting some ways in which images can instill falsity (exemplified by advertising images). Their general concern is not with logic but with other senses of truth and falsity. These are highly relevant when trying to assess and understand individual cases of imagery, but have less applicability when dealing with a collection especially as an archivist or cataloguer who is unable to know all the different questions which will be posed; unable to predict the future uses of the collection and hence how it can or should be ‘summarized’ in catalogues entries.

Practical Approaches

If purely visual propositions are not possible, what practical options are available to those who have to deal with a large picture or video collection? The answers to this question provoke a fresh return to some of the older theoretical positions concerning the relationship of word and image. Is access to imagery (still or moving) essentially mediated by language? Our inability to make a non-verbal index to an image collection suggest that our reliance on words is unavoidable. Later we will consider what might be a video-summary of a ritual. One general term introduced by a pioneer

of digital video libraries is that of 'the surrogate': a composite which stands-in for the original (O'Connor 1985). We shall consider two separate types of surrogate: filmstrips and salient stills. The basic idea of a filmstrip is based on contact prints (Christel, Winkler & Taylor 1997). The principle is that adjacent frames may be considered simultaneously so that images usually seen sequentially (distributed in time) are viewed simultaneously (but distributed in space) which is impossible when watching a film/video. Since in most clips there are too many frames in a video to print or display them all, the video must be sampled, for example, displaying one frame per clip, or one frame per minute. The result is an interesting abbreviated document.[1] One point of interest is how little gets lost. Preserving the sound track means that the chronology can be preserved, so one can use the partial representation of the video as an index to the entire video as shot.

A related idea is that of the 'video skim' in which the still image of the filmstrip is replaced by a small piece of video, so that some of the dynamism of the original is preserved. A very simple form of video skim can be generated by sampling the first ten seconds of every minute (or five minutes etc.) of original video and splicing them together. The resulting document is considerably shorter than the original and reflects its content in a systematic and comprehensible manner. However, neither filmstrip or video skim really approach the self-referential logical properties that make language so powerful and concise (see below). These are promising

---

[1] A similar approach was developed for practical reasons by Michael Fischer when he started using digital video in Cook Islands in the late 1980s see <http://lucy.ukc.ac.uk/tradition.html> (accessed June 8, 2004) for an example.

beginnings, and they mark a small but significant step in the development of ways of dealing with visual material that begin to approach the complexity of language.

Salient stills

Salient stills were invented and developed by a team of researchers in MIT in the early 1990s (Teodosio and Bender 1993, Massey and Bender 1996). As yet they have not been widely used or distributed. The technology for preparing them is proprietary, which may be an obstacle to their further development as an indexing tool, as suggested below. The salient still is a single image which represents a segment of film or video in a systematic and intelligible form. Consider a video clip in which the camera pans round a room then focuses on an individual before zooming in on their face. A salient still based on this clip would contain a low quality image of the room with the individual shown in greater detail, and their face in still greater detail. Essentially the salient still exploits (and represents) the redundancy of information contained in video footage, by changing the resolution (hence the image quality) of portions of the salient still image which represents the clip. A zoom is a clear example of this redundancy between multiple film frames: there is nothing in the final frame that was not in the initial frame, but what does appear is shown at greater magnification. The salient still translates this into image quality, crudely into changes of pixel resolution. Thus the face on which the camera has zoomed may end up at 300 dots per inch (henceforth d.p.i.) while the surrounding room is represented only at 72 d.p.i., the individual's body being an intermediate resolution of 120 d.p.i. and so on. A salient still is like a painting in which the background has been only lightly sketched, most of the artist's attention being given to the foreground subject. Salient stills can function as an indexing device by automatically identifying the foreground

and background of a video clip and representing an entire clip by one salient still. The results are powerful and intriguing. The limited number of tests reported in the literature suggest that users have no problem interpreting them, although, unsurprisingly, there has as yet been no cross-cultural testing. The algorithms that produce a salient still from a video clip have been patented and can be licensed.

IMAGE 1. How a zoom becomes a salient still. Source Teodosio and Bender 1993 (used with permission)

Intriguing possibilities result from a combination of the salient still and the filmstrip: a salient still could be used to represent each clip and the sequence of clips displayed as a sort of filmstrip to represent a summary account or surrogate of an entire video.

I note that there is no easy parallel to salient stills (or filmstrips) for sound recordings. For those who want to create an index of sound recordings the only option available seems to be some form of sampling. This could be linked to filmstrips: so, for example, the first, middle, and last ten seconds of sound from a film clip could accompany the filmstrip that indexes its visual content. Whether this is a better strategy than a random sample, and how long the sample segments must be, are empirical questions which call for further research. The CMU Informedia project created audio skims on a parallel with video skims. First, sound tracks are shortened by eliminating repetitions and reducing pauses. Then salient phrases in a sound track were identified using automatic sound processing. A parser identified noun-parts which were passed to a speech to text recognition program. The result was a written

text which could be analyzed with textual tools and used to select salient parts of the sound track, which themselves become the basis for selecting the extracts or video skims, so connecting video and audio skims. However, the move from the audio to the textual as the necessary preliminary step for identifying salient parts of the sound track is, for my current purposes, evidence of the weakness of audio-only analysis.[2] Language has once again become the essential means of accessing the visual.

Theoretical Implications

These examples put High Theory in a curious position. They both illustrate and circumvent the force of the suggestion that a photographic image 'is a message without a code' (Barthes 1977:17). By starting with the seemingly prosaic problems of running and searching photographic archives we are led to view theory in a different way. Archives pose challenges both for the archivists who have to manage the collections and for the users who want to call up material they have never seen. This returns us to Barthes and Benjamin's ideas of imagery as embedded in language. A purely visual, an *entirely* non-verbal, catalogue cannot be made to work. Work on face recognition provides an ironic confirmation of this: the major goal of the developers of face recognition and other biometric systems is to attach names to faces. This resembles the way that captions are attached to paintings and photographs in galleries, for all the far more serious (sinister according to some) purposes behind

---

[2] Some experimental audio- based systems have been developed such as the Hewlett-Packard Research lab's 'SpeechBot'. See <http://www.hpl.hp.com/news/2003/apr_jun/SpeechBot.html> (accessed June 8, 2004), see also Takasu, Yanase, Kanazawa & Adachi 1999 for another example.

it. Image recognition combined with databases of individuals, as maintained by national governments and international organizations such as Interpol, offer the possibility of realizing Orwell's dystopic vision of 1984 only a few decades late. The panopticon gains teeth when those at its heart, those watching, can easily discover *who* they are looking at …

Indexing in Real Time

Life is a rich, complex multi-media experience. Any representation, in any form or medium, is necessarily incomplete. The challenge is to be clear about the manner and type of partiality, the degree of incompleteness. It is more important to be able to give an account of this in some subjects or disciplines than in others. Thus we find a continuum from hard science to the arts (opinions vary about where anthropology falls on this continuum but that is a different issue).

Consider as an example, a major sporting event or a complex royal ritual (for example, a funeral or a coronation). I have studied one such ritual that takes place every two years and which lasts more than seventy-two hours. (I note that it took twenty words to say as much). To demonstrate the ritual by, for example, actually staging a performance would, if done accurately, take at least seventy-two hours. That would be an interesting and complicated type of representation, preserving many of the temporal relationships but losing the specificity of time and place of performance, as well as the identities of the performers. Unless, that is, it could be arranged that the usual performers 'stage' a performance in the usual place, resembling a rehearsal (or a performance specially undertaken for film cameras, tourists etc). Although different issues of representation arise when people play themselves, such issues are nonetheless real. Current examples of this include the teaching children local traditions in schools, or shows performed either to teach other groups about local

traditions (examples can be found in North America (e.g. Simonelli 1996; Cleary and Peacock 1998) and Australia (e.g. a CD designed for use in Australian schools, Glowczewski 2001)) or to impress visiting dignitaries. They are all types of representation, and are partial and incomplete in a variety of ways, both by design and by happenstance. As long we are clear about what is going on such incompleteness is not necessarily a bad thing.

Anthropological (or any other variety of) description is necessarily incomplete. A few words are made to represent a host of complex and varied activities, for example:

> The dance continued for three days with increasing disorder as the amount of alcohol consumed took effect.

A verbal account says an immense amount (and omits even more). Language, whether spoken or written, has these attributes because it admits of degrees of abstraction which images (such as those in photographs and film), arguably, do not have, or have not had until recently through techniques considered above, such as the salient still.

Consider non-verbal ways of documenting the hypothetical dance described above: a set of photographs might have been made by the ethnographer which could demonstrate the point. Some annotation would be needed to represent the three days' duration. Time lapse photography, for example, taking a photograph an hour from the same vantage point, could also be used, leaving the viewers to work out that the 72 photos were taken over a three-day period. What is true for photographs is also true of video or film.

The problem is that neither photographs, film/video, nor sound recordings have the properties of encapsulated summary that is trivial to undertake in words. Some of the logical features of language have no easy correlates in visual

representations. For example, language contains statements such as 'All Cretans are liars' and with quantification (to use the term from formal logic) comes the possibility of negation, summary, abbreviation and concision (see above and Worth 1981 for an extended discussion of this point). Perhaps this may help to explain why so few essays in the visual anthropological literature are predominantly wordless; why they have to use words to discuss the visual. MacDougall discusses the visual (specifically anthropological film but the points he makes are general) and its relationship to anthropological knowledge (1998 see especially 73ff). He treads a delicate path between evocation and display; between Russell's knowledge by acquaintance and knowledge by description (which he uses 1998:77-8). In his vision 'filmic ethnography' is achievable through audience engagement with the filmmaker's work : 'filmic ethnography … tends to draw attention to *relations*. Writing can give theoretical causal explanations, but a film can only suggest causal relationships within a given context' (1998:75). It may be satisfactory to leave it to the audience of a finished film to deal with the suggestions made therein but such an approach cannot help the makers or users of visual archives.

Natural languages can be paraphrased and analyzed (also negated), exhibiting aspects of formal logic as just mentioned. Paraphrase allows for concision, and summary; this is what the visual and audio channels lack. Although films may be paraphrased by a human 'editing down' to produce a synopsis (explicitly paralleling the editorial shortening of *Reader's Digest* shortened books) and music can be summarized in terms of its 'themes', neither are possible for the unstructured video or audio streams from raw 'point and shoot' recordings. Consider Warhol's film of the Empire State Building (1964) where part of the point is that the film lasts eight hours; or, for example, parallels outside artistic cinema: surveillance cameras and web cams.

We should note that edited summaries such as the (misleadingly titled) *Big Brother - Uncut* (2000) compilations are not viable counter-examples because they concentrate on the most exciting and salacious parts of the footage. They do not give a fair representation of the sheer tedium (for viewers and inmates alike) of life in the *Big Brother* house.

Related to the problem of paraphrase and summary in visual material is another similar, problem of indexation: there is no obvious visual parallel to a book index, due to the problems discussed above. How can one summarize five minutes of video and encapsulate it in one pithy, evocative 'phrase' as happens routinely in verbal media e.g. an index entry which reads 'indexing pp. 5-8'? Visual material poses very clearly the twin archival challenges of 'access' and 'retrieval' the users want to find *all* the relevant material and *none* of the irrelevant material (Blair and Maron 1985). Even specialized bibliographic services such as Anthropological Index Online suffer such problems. They particularly afflict general indexing services for the internet such as Google. On 30th October 2004 I searched Google for material relevant to this paper. I was offered 117,000 results for "indexing digital video library". There is no way of telling which of these is *actually* relevant to this paper; I do not have time to check them all. Therefore there is also no way of being certain that I have not missed a vital source.

Some early experiments on visual material are worth noting. Birkbeck College London had a collection of photographs of classical paintings (digitized as part of the Morelli project: see Vaughan 1992). These were in black and white so a vast amount of information had already been discarded, as had information about medium (watercolor, oil, tempera or fresco) and, for example, the thickness of the paint. The 10,000 or so photographs in question were digitized and then systematically reduced

to a grid of 4 by 4 pixels giving a total number of 65,536 possible combinations. The claim has been made that, even with b/w pixels, such a grid can help distinguish works by the same painter from those by others. In the pilot system, the user selected a photograph of one painting and could then retrieve other photographs with the same basic 4*4 template. Other possibilities existed, for example looking for sharp gradients, which possibly would select a distinctive style such as chiaroscuro, and so on. Although demonstrated in the early 1990s, such systems have not (yet) become widespread. By the time of writing some services have become available which realize some of these early visions: examples include the use of IBM's 'Query By Image Content' system to search the St Petersburg Hermitage Museum catalogue[3], James Wang's SIMPLIcity (Semantics-sensitive Integrated Matching for Picture Libraries) system[4] and the AHDS Visual Arts Collections service[5].

Visual anthropology and the absence of a meta-language

These issues raise the question of what representations are. We must also ask how different styles of representation have different properties which constrain what are and can be done with them, independent of what is being represented. As Meyer put it:

---

[3] <http://www.hermitagemuseum.org/fcgi-bin/db2www/qbicSearch.mac/qbic?selLang=English> (accessed June 8, 2004).

[4] <http://wang16.ist.psu.edu/cgi-bin/zwang/regionsearch_amico.cgi> (accessed June 8, 2004).

[5] <http://www.ahds.ac.uk/visualarts/collections/index.htm> (accessed February 27, 2005).

> David Marr was fascinated and intrigued by the complex relations that exist between the choice of a representation of a signal and the nature of the operations or transformations that such a representation permits. He wrote […] 'A representation is a formal system for making explicit certain entities or types of information, together with a specification of how the system does this. And I shall call the result of using a representation to describe a given entity a description of the entity in that representation.
>
> Thus, there is a trade off: *any particular representation makes certain information explicit at the expense of information that is pushed into the background and may be quite hard to recover*'. (1993:9-10 his emphases)

A much quoted example of this is the difference between Arabic and Roman numeral systems. Both are equally good at representing numbers as bare quantities but multiplication and division are difficult in the Roman system.

Another approach (Krüger 2001) distinguishes feature-based from template-based representations. In the former an image is decomposed into a set of constituent features, and is then represented as the sum of its parts, each of which may have a compact and relatively simple representation. In the latter an image is identified as being a kind of X, and then compared to a stereotypical X, noting the differences between the token and the stereotype.[6] In principle, the image can be reconstructed from the record of those differences, which can then be used to 'distort' the stereotype to recreate the original image. As Krüger (2001) points out, is it possible to do both together, as exemplified by Gabor Wavelet Networks, the basis for holograms. A hologram is a record of the interference pattern of two beams of coherent light after they have been shone onto an object. The appearance of that object can be subsequently re-created from the resultant interference pattern. The hologram represents its object in ways similar to, but more complicated than, the ways in which photographs represent their objects. There are physical processes involved but, as a

[6] Thereby instrumentalising Galton's idea of an average face.

large amount of theory has shown, there is more to it than 'only' physical interaction. The question here is whether or not the 'something else' itself is understood as being visual. The case of a hologram makes the point since in strict terms it is a mathematical representation of a 'visible signal' as a weighted sum of constituent waves.

Some of these issues are also discussed in the psychological literature, by e.g. Stephen Kosslyn (1980, 1994) and in discussion of the meaning of analogical representations (Hofstadter 2001). These approaches to representation are wholly different from the way it is usually discussed in the art historical, philosophical, and the anthropological literatures. The profound differences and apparent incommensurability of these approaches were part of the impetus for this paper. Yet having introduced these dilemmas I propose to leave the more abstract discussion and use the problems of concision and indexing to anchor my consideration of these issues.

Concision is an interesting concept and one that is difficult to make precise. What is it exactly that one is talking about, especially when applied to visual media? In feature film terms one could contrast *Heaven's Gate* (219 min.), or Abel Gance's *Napoleon* (240 or 300 min), with *Un Chien Andalou* (16 min.) to make the point. Only the latter is concise, or is it merely short? The two are not synonymous. Surrealism helps make the point: '*Ceci n'est pas un pipe'*.

IMAGE 2 René Magritte '*La trahison des images: Ceci n'est pas une pipe*' 1929 *The treachery of images (This is Not a Pipe)* Los Angeles County Museum of Art, Purchased with funds provided by the Mr. and Mrs. William Preston Harrison

Collection. (c) 2004 Museum Associates / LACMA / ADAGP, Paris and DACS, London

As well as succinctly exemplifying the difference between objects and their representation (see Foucault 1983), this image also raises the type-token distinction and the logical idea of quantification on which it rests. You cannot see the concept 'pipe' you can only see a particular pipe. We encounter tokens, never types, hence we cannot easily, in visual media, make general statements, let alone arguments (for all of John Berger's claim to the contrary in *Ways of Seeing* (1972) which ends with a purely pictorial essay; he has had few successors, and very few in visual anthropology.[7]

Despite the possible counter-example of minimalism,[8] it seems that the idea of a concise image may be a contradiction in terms. This is challenging: is there no possibility of 'concise representational images'? Is it self-contradictory to claim that there can be images that represent the world yet are concise? For example, pencil sketches, or black and white photographs are concise in that they have omitted color

---

[7] Chris Marker's filmworks e.g. *La Jetée* 1962 (released 1964) and *Sans Soleil* 1982, are primarily sequences of still images but the contrasts and tensions between the words of soundtrack and image are critical. Godfrey Reggio's work (e.g. *Powaqqatsi*, 1998; *Koyaanisqatsi* 2003) is more purely visual but his films make arguments only in the wider sense of the term.

[8] Exemplified by the work of artists such as Alan Chalton and Robert Rauschenberg, and Malevich's painting of a white square on a white background.

information. An architectural drawing omits shade and tone, as do some line drawings. Yet to claim that these as concise or that they summarize in the same way as a published abstract summarizes (hence 'represents') a journal article feels like a category mistake.

Without wishing to retrace the entire history of art with regard to representation and abstraction, I consider only video images which, of course, have a complex, mediated, indirect relationship to objects and events in the world. Any representation we see on the screen is constrained (note: not determined) by the laws of optics and by the world technological economic complex, quite apart from the decision I make about when to press the record button, how to zoom, pan and so forth. As we have seen in the discussion of filmstrips and salient stills, above, there are now ways of systematically summarizing video.

Abstracting visual data

In order to connect back to the theoretical discussions of Barthes and Benjamin I now return to a theme mentioned in relation to audio summaries above. I want to consider ways of abstractly representing visual information in a *non-textual* fashion, (as used by the so-called *Content Based Information Retrieval* systems).

The idea of an abstraction of a visual image is best illustrated by some examples. Consider the following image which is a grey-scale image derived from a color original. The black and white version has discarded all the color information (and is correspondingly some four times smaller in terms of file size). Despite the reduction in information, its anthropological use seems scarcely affected: the individuals in both the original color and reduced black and white photographs are

similarly identifiable. However, information about e.g. clothing and objects has undoubtedly been lost. Such objects are often the subject of anthropological enquiry.[9]

IMAGE 3 A field photograph

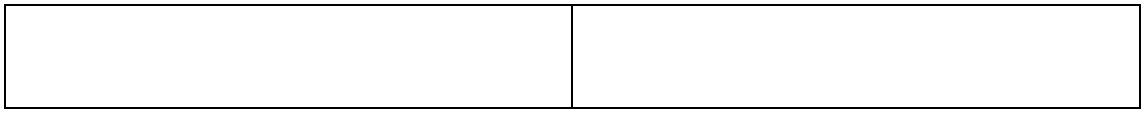

IMAGE 4 Photoshop 'Levels' Representation of Image 3 (left)

It cannot be said that Image 4 has an established anthropological use. This shows a representation of Image 3 as presented via the 'Levels command' from an image editing program such as Adobe Photoshop. This displays a graph showing the distribution of different pixels at their various degrees of grey. Each pixel has a particular color value (singly for a grayscale image, repeated for each of the color components of a color image). The graph shows the number of pixels of each different value of grey (red, green, blue, etc). As a consequence, two images with a

[9] In another context Woody (1998) discusses the introduction of color printing in postcard manufacture. Both he and other authors in the Geary and Webb (1998) collection provide examples of black and white and color versions of "the same" image. However, in these early examples color was being added by the printers to a black and white negative rather than being removed from a color original (see figs. 28-33, pp 37-39, figs. 117 and 188, p.158 and figs. 78 and 79, p 117).

similar shape graph, irrespective of their representational content, will have a similar range of tonal values. The possibility of searching and classifying on purely pictorial qualities now arises. In such a system a color image could be represented as a vector in three dimensional 'color space' which provides a graphical representation of the distribution of tones across the three primary colors.

IMAGE 5 Dimensions in 'color space'

A practical realization of a search and retrieval system that works along these lines may be found in the online 'ZOMAX' system which provides 'Color Image Invariant Segmentation and Retrieval'. [10] Presented with a sample image, the Zomax system finds other images with similar colors. Related work is being done in the potentially lucrative fields of face recognition and pornographic image detection, the former for a variety of security reasons, the latter in order to improve child protection filters. However, what is lost in such representations is the possibility of recognizing individuals. The system may be able to say 'there are humans in these images' but not see them in the same way as humans see the black and white image in Image 3: as representations of identifiable individuals (see comments on face recognition above).

---

[10] See Gevers and Smeulders 1992 and 1999 for background and <http://zomax.wins.uva.nl:5345/zomax/HTML/zomax.html> (last accessed June 8, 2004) for online demonstrations.

Another possibility arises when one decomposes the image by radically reducing the number of pixels (see above). The images above are made up of grids of several hundreds of pixels in each dimension (in the original scan 1464x2176 pixels).

Reducing the image to 11x16 pixels one obtains the following images.

IMAGE 6 Reduced resolution images: 11x16 pixels



Or consider a similar operation on Magritte's image

Despite the huge loss of information I hope readers will agree that they retain some information about the look and composition of the originals. In that case they are, in some sense, summaries of these originals.

Further reduction to 8 x 8 pixels is also possible, and the resulting index of the photograph could also be used as a retrieval tool: one could draw in a simple grid and thereby find all images with a strong diagonal composition, or those with a dark foreground. These images display their connection to the originals more comprehensibly to humans than the abstraction of the color space, the mathematical representation of a hologram or the 'levels' display in editing software (see above).

Conclusions

I draw two conclusions from this, one theoretical, one more practical. At a practical level it should be noted that some pilot versions of systems have been developed using both filmstrips or skims and more abstracted indexing varieties. The intriguing

thing about the latter is how unsatisfactory the resulting systems are for users (see the user studies reported in Christel et al. 1997, 1998, 1999, 2000 and Lee & Smeaton 2002). It appears that the very strength (representing the visual in visual modes) which they gain by being disconnected from textual information about content is also their weakness. I suggest that systems usable in the real world must employ features of all the different types of indexing that have been briefly considered here (see, for example, Smith and Kanade1997; Lee & Smeaton 2002). As was said above, one intriguing possibility would be to combine salient stills with filmstrips.

The theoretical point was alluded to in the introduction: visual representations are not language-like. They lack the logical properties of quantification implicit in all natural languages: it is impossible to provide a visual equivalent to the statement 'all painters paint'. This has profound implications for how we understand and discuss any system of visual communication. Indeed it returns us to some of the concerns of photographic theory (see Benjamin 1980 (1936); Barthes 1977:36ff and more recently, Mitchell 1994; and Scott 1999). So, contentiously, we apprehend images only through a textual mesh. 'Pure' or unmediated apperception of images cannot occur. Work in image theory has pursued some of the implications of such positions, but here I have examined some of the practical implications for those attempting to manage and search image collections.

The description of an image characterizes the non-verbal in verbal terms and the description is then open to summary and concise simplification. It is too easy to overlook the first step that took us away from the visual object to a linguistic representation, a truly 'radical translation'. The temptation is to employ the techniques of analysis developed for literary and textual subjects, but these are often

inappropriate. In short: if we cannot index images in ways similar to indexing texts then we cannot analyze them as if we were dealing with texts.

Bibliography

Barthes, Roland

1977 Image-Music-Text. London: Fontana-Collins.

Benjamin, Walter

1980 [1936] The Work of Art in the Age of Mechanical Reproduction. *In* Classic essays on photography. A. Trachtenberg, ed. Pp. 199-216. New Haven, Conn: Leete's Island Books.

Berger, John

1972 Ways of Seeing. Harmondsworth: Penguin.

Blair, David C., and M.E. Maron

1985 An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System. Communications of the Association for Computer Machinery 28(3):289-299.

Christel, M., A. Olligschlaeger, and C. Huang

2000 Interactive Maps for a Digital Video Library IEEE Multimedia 7(1), 60-67.

Christel, M., D. Winkler, and C. Taylor

1997 Multimedia Abstractions for a Digital Video Library. *In* Proceedings of the 2nd ACM International Conference on Digital Libraries. Pp. 21-29. Philadelphia, PA. July. Also available as an electronic document http://www.informedia.cs.cmu.edu/documents/Acm_dl.pdf, accessed February 24, 2005.

Christel, M., D. Winkler, R. Taylor, and M. Smith

1998 Evolving Video Skims into Useful Multimedia Abstractions *In* Conference on Human Factors in Computing System (CHI'98). C. Karat, A. Lund, J. Coutaz, and J. Karat, eds. Pp. 171-178. Also available as an electronic document http://portal.acm.org/ft_gateway.cfm?id=274670&type=pdf&coll=GUIDE&dl=GUIDE&CFID=40077699&CFTOKEN=67127030, accessed February 24, 2005.

Christel, M., A. Warmack, A. Hauptmann, and S. Crosby

1999 *Adjustable Filmstrips and Skims as Abstractions for a Digital Video Library In* IEEE Advances in Digital Libraries Conference 1999, Baltimore, MD. Pp. 98-104. Also available as an electronic document http://www.informedia.cs.cmu.edu/documents/adl99.pdf, accessed February 24, 2005.

Cleary, L. M., and T. D. Peacock

1998 Collected wisdom: American Indian education. Needham Heights, MA: Allyn and Bacon.

Foucault, Michel

1983 This is Not a Pipe. trans. James Harkness. Berkeley: University of California Press.

Gevers, Theo and Arnold W. M. Smeulders.

1992 Indexing of Images by Pictorial Information. *In* Visual Database Systems, II. Proceedings of the IFIP TC2/WG 2.6 Second Working Conference on Visual Database Systems. Budapest, Hungary, 30 September - 3 October 1992. IFIP Transactions A-7 Elöd Knuth, Lutz Michael Wegner, eds. Pp. 93-100. North-Holland: IFIP. Also available as an electronic document http://staff.science.uva.nl/~gevers/pub/paper_budapest.ps, accessed February 24, 2005.

Gevers, Theo and Arnold W. M. Smeulders.

1999 Color based object recognition. Pattern recognition, 32(3): 453-464.

Glowczewski, Barbara

2001 Dream Trackers - Yapa Art and Knowledge of the Australian Desert (CD Rom). Paris: UNESCO Publishing.

Gombrich, Ernst H.

1960 Art and Illusion. A study in the psychology of pictorial representation. London: Phaidon.

Hofstadter, Douglas R.

2001 Analogy as the Core of Cognition. *In* The Analogical Mind: Perspectives from Cognitive Science, Dedre Gentner, Keith J. Holyoak, and Boicho N. Kokinov, eds. Pp. 499-538. Cambridge: The MIT Press/Bradford Books.

Kosslyn, Stephen M.

1980 Image and Mind. Cambridge: Harvard University Press.

1994 Image and Brain: The Resolution of the Imagery Debate. Cambridge, MA: MIT Press.

Krüger, V.

2001 Gabor Wavelet Networks for Object Representation. Technical Report Number 4245, University of Maryland.

Lee, Hyowon and Alan F. Smeaton.

2002. Designing the User Interface for the Físchlár Digital Video Library. Journal of Digital information 2(4). Online resource http://jodi.ecs.soton.ac.uk/Articles/v02/i04/Lee/, accessed 20 March 2003.

MacDougall, David

1998. Transcultural Cinema. Princeton: Princeton University Press.

Massey, Michael and Walter Bender.

1996. Salient stills: Process and practice. IBM Systems Journal 35:557-73.

Meyer, Y.

1993 Wavelets: Algorithms and Applications. Philadelphia: Society for Industrial and Applied Mathematics.

Mitchell, William J.

1994 How to Do Things with Pictures. *In* The Reconfigured Eye: Visual Truth in the Post-Photographic Era. Pp. 190-223. Cambridge: MIT Press.

O'Connor, Brian C.

1985 Access to Moving Image Documents: Background Concepts and Proposals For Surrogates for Films. Journal of Documentation. 41:209-220.

Roskill, Mark W., and David Carrier

1983 Truth and Falsehood in Visual Images. Amherst: University of Massachusetts Press.

Peacocke, Christopher

1987 Depiction. The Philosophical Review 96:383-410.

Scott, Clive

1999 The spoken image: photography and language. London: Reaktion Books.

Simonelli, Richard

1996 Partnering with Indigenous Education. On Common Ground 7. Online resource http://www.yale.edu/ynhti/pubs/A20/simonelli.html, accessed October 30, 2004.

Smith, Michael A. and Kanade, Takeo

1997 Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques. *In* Computer Vision and Pattern Recognition Conference. Pp. 775-781. San Juan, Puerto Rico.

Takasu, Atsuhiro, Takashi Yanase, Teruhito Kanazawa, and Jun Adachi

1999 Music Structure Analysis and Its Application to Theme Phrase Extraction. *In* Proceedings of the Third Annual Conference on Research and Advanced Technology for Digital Libraries, Paris. Lecture Notes in Computer Science 1696 Abiteboul, Serge and Anne-Marie Vercoustre eds. Pp. 92 – 105. New York: Springer.

Teodosio, Laura & Walter Bender.

1993. Salient video stills: content and context preserved. *In* Proceedings of the First ACM International Conference on Multimedia '93, Anaheim, CA, 1993. Vol. 6. Pp. 39-46. ACM Press. Also available as an electronic document http://portal.acm.org/ft_gateway.cfm?id=166270&type=pdf&coll=GUIDE&dl=GUIDE&CFID=38764271&CFTOKEN=28655098, accessed February 24, 2005.

Vaughan, William

1992. Automated Picture referencing: a further look at 'Morelli'. Computers and the History of Art 2:7-18.

Worth, Sol

1981 Pictures Can't Say Ain't. *In* Studying Visual Communication. L. Gross. ed. Pp 162-83 Philadelphia: University of Pennsylvania Press. Online resource http://astro.temple.edu/~ruby/wava/worth/ssev.html, accessed May 24, 2004.

Zeitlyn, David

1996 Regarding machines for the suppression of time. Online resource http://lucy.ukc.ac.uk/dz/it-va/it.html, accessed October 30, 2004.

Films

2000 Big Brother - Uncut - The Official Video. 146 min.: Warner Vision International DVD

Buñuel, Luis, dir.

1929 Un chien andalou 16 min.

Cimino, Michael, dir.

1980 Heaven's Gate. 219 min. United Artists. Holywood.

Gance, Abel, dir.

1927, restored version 1981 Napoleon (240 or 300 min. depending on version). Zoetrope Studio. Holywood.

Marker, Chris , dir.

1962 (released 1964) 28 min. La Jetée

1982 Sans Soleil 110 min. New Yorker Films. New York.

Reggio, Godfrey, dir.

1998 Powaqqatsi. 99 min. Institute For Regional Education. Santa Fe.

2003 Koyaanisqatsi. 87 min. Institute For Regional Education. Santa Fe.

Warhol Andy, dir.

1964 Empire. 485 min. Andy Warhol. New York.