

A VIEW FROM THE ARCHIVE

PETER ENSER
University of Brighton

P.G.B.Enser@brighton.ac.uk

Image retrieval, like information retrieval generally, is not an operation which owes its existence to the invention of the computer. On the contrary the curation of visual heritage has a very long tradition which relies on the application of the human intellect and domain knowledge to appreciate the semantic content of still and moving images such that the knowledge encoded in them can be recorded and made available in response to need.

The computer science and vision research community have spent years and a great deal of money trying to dodge round the uncomfortable fact that, in general, image material is sought for its semantic content, and that the great majority of image retrieval transactions necessarily depend on textual metadata and verbal queries.

Jørgensen has expressed it neatly: “the emphasis in the computer science literature has been largely on what is computationally possible, and not on discovering whether essential generic visual primitives can in fact facilitate image retrieval in ‘real-world’ applications.” [Jørgensen, C. *Image retrieval: theory and research*, p.197].

My message to the computing image research community is: please come out of your labs and learn from the experience of picture librarians and archivists, and pay more heed to what clients of image collections actually want.

Studies of user needs have shown clearly the high incidence of a number of types of request for which not only is textual metadata essential, but their resolution demands a high level of conceptualisation little informed by low level features.

The first such type of request concerns the client’s desire to recover images of specific features *identified* by proper name.

No matter how sophisticated shape-matching procedures may become in future – and nobody looks forward to that more than me – there will have to be a defining textual tag somewhere.

This leads to a second, pressing need for supporting textual metadata. This is when the *significance* of some visual feature is at issue. Significance is an attribute which is unlikely to have any visual presence in the image; it often reflects the reason for the image having been created in the first place. Significance frequently takes the form of the first or last occasion when some visible feature occurred in time, or the first/only/last instantiation of some physical object.

There is a third category of requests, for which the recovery of relevant images demands semantic reasoning of a high order. At issue here are those situations where the attempt is being made to capture or represent a culturally specific concept. Often the client wants to use the image as a visual metaphor or metonym – a high-level semiotic device.

A related and frequently encountered requirement is to trigger an emotive response in the viewer. We note in passing the power of the caption-writer to provide the contextual anchorage which enforces a particular interpretation to be placed on a depicted scene or feature.

The final case to be considered are those instances where the client names features which must *not* be present in the retrieved image. Provision is sometimes made in keywording schemes for the logical NOT operator to be employed, but where the attributes in question are not specified as keywords, resolution of this type of request can only be achieved by human visual inspection. A search strategy which did not first approach these needs via textual metadata is difficult to contemplate.

None of this is to deny that there are significant applications for which verbalisation of either image content or need is inappropriate or impossible. We know little enough about how semantic content is realised in these instances.

My conclusion is that the semantic gap is really a semantic chasm; coming to terms with this feature in the image retrieval landscape will demand a much closer working relationship between the computer scientists and the image/film/video archivists than has been our experience to date.

One approach which holds out some promise is ontologically annotated image sets, and cross-paradigm metadata. With my colleagues Criss Sandom and Paul Lewis I am currently engaged upon an AHRB-funded project which is taking such an approach, informed by a plurality of user communities and their varied expressions of need for visually encoded knowledge.