9 THE VISUAL THESAURUS: A PRACTICAL APPLICATION

Corinne Jörgensen
Centre for Science and Technology 4-290
Syracuse University
USA
and
Peter Jörgensen
Interactive Publishing
Hamilton, USA
and
Matt Hogan
Centre for Science and Technology
Syracuse University
USA

Image storage and retrieval are currently the focus of much research, yet an understanding of such basic issues as the relationships between text and image, the effects of human processing of images, and the impact that these could have on systems design remains hazy. Many of those concerned with image processing and recognition rely on technological advances in such areas as machine vision to solve problems of retrieval, remains focused on hardware and software advances.

Other researchers are concerned with issues of intellectual access, and various text-based mechanisms such as thesauri and indexes have been proposed and implemented. These two approaches, however, are similar in that both are what could be termed "system-driven;" both begin from the standpoint of using existing systems or systems methodologies, whether they be text-based systems of classification or algorithmic manipulations of data, rather than from a theoretical basis and understanding of both image-text relationships and human interaction with images. Therefore, image storage and retrieval systems are being designed without an understanding of what system functionalities are really needed by the human user in image retrieval.

Currently, image retrieval remains primarily text-based. The problems of having information in a visual mode and needing to find the written language to describe the object, name the item, and explain its functions and its relationships with other objects is widely recognised (Hogan, Jörgensen, and Jörgensen, 1991). Aside from these difficulties, we maintain that treating visual objects in a purely textual way is too limiting. Tools need to be made available to enhance exploration, reveal unexpected relationships, and stimulate the creativity of researchers. We do not propose to throw out textual methods of dealing with visual materials; we would like to augment these methods with others which we believe have been neglected for too long. We believe that research is needed to explore

the added functionalities which visual retrieval could add to information retrieval. Information retrieval has been characterised too narrowly by systems developers, who, by choosing efficiency as a primary goal, are ignoring a host of other needs of researchers.

The image-text dichotomy

There is a basic assumption surrounding the relationship between images and texts which has led to their specialised treatment in information retrieval systems. Images are pictorial and non-verbal, while texts are verbal. While there may indeed be a true image-text dichotomy, we would argue that the nature of the relationship is not well-enough understood to design image storage and retrieval systems around this assumption. Human information seeking in texts and images must be better understood before truly usable systems can be designed. The nature of human image processing and retention, which suggests that details are less important than an overall schema, has implications for the amount of detail necessary to enable searching. Images are also retained in a highly structured manner; this would suggest that textual indexing assigned to an image should maintain these structured relationships, rather than simply being assigned in a linear manner.

By developing a practical application of a "visual thesaurus," a system which provides flexible search capabilities yet maintains authority control, we intend to explore this assumption and develop some new approaches to image indexing and retrieval which incorporate spatial and iconic representations.

Methodology

In order to seek answers to the above questions we adopted a case study approach, analysing the information and image retrieval needs of one researcher in a narrow field (Classical Archaeology). This approach enables us to achieve better control over the large number of variables inherent in such exploratory research. The researcher was interviewed to determine her specific information seeking and retrieval needs within the context of an image/text database, and a prototype system was created. Research into information seeking within the image database and evaluation of the system consists of taped sessions of human-system interaction, with the user being asked to talk aloud as he/she uses the system. The users are both the expert researcher and her students who are just gaining familiarity with the field. Additionally, a small group of information science researchers will interact with and evaluate the system.

Context of the problem

Within the general discipline of Classical Archaeology, the specific images of concern are terra cotta objects which served as votive offerings in temples at particular sites during a period of time from 600 BC to the first century AD, spanning the Archaic, Classical, and Hellenistic periods. There are a large number of figurines (both human and animal) which vary in pose, dress and hairstyle, objects held, and materials and manufacturing techniques (mold cast, hand-modelled, etc.) Within specific mold series there are also differences. Many of these figurines are not complete.

There are three major sources of figurines, and therefore of images which need to be collected, catalogued, and retrieved:

- raw material directly from excavation sites;

- museum collections;
- library collections of printed books or catalogues of museums, which contain large number of photographic plates of the objects.

The researcher is initially faced with the problem of organising and cataloguing these figurines in a meaningful and useful way. Materials may have been previously organised by some other system, but that system may not be "trustworthy" enough for the researcher.

By studying the physical evolution of the figurines the researcher hopes to gain an understanding of religious, social, and political practice in Greece and Rome. In order to have enough data to be able to make meaningful generalisations and substantive conclusion, as many images of figurines and their associated textual records need to be easily accessible. Much of the initial work of classifying these objects is done on the basis of physical, visual characteristics, and much of the theoretical work derives form comparisons of groups of objects and their specific characteristics.

Therefore the researcher needs a system which both facilitates the initial organisation and cataloguing of these images and retrieves specific types of images or groups of images. The researcher needs to search for both prototypes (standing Venus figurine) and specific image characteristics (Venus figurine wearing Rosette earrings).

System development

Several concerns guided system development. The researchers in this field need the ability to take the system into the field, the museum, or the library in order to input data or to retrieve images for comparison and research. Therefore, the system needs to be easily portable and have satisfactory graphics capabilities. We chose to develop the system using HyperCard on the Macintosh for several reasons. An image-intensive database such as this requires a system that can handle large numbers of image files easily. The availability of QuickTime image compression technology allows us to store a 9cm by 15cm by 8 bit (232x432x8) image in a 12KB file - a five times savings over the non-compressed file. The graphic display capabilities of the Macintosh allow these images to be displayed on any model that supports QuickTime, regardless of the bit depth of the screen. The complete compatibility between models allows us to develop the system on a high-end machine (Quadra 700) with full confidence that it will work on smaller systems (PowerBooks and LCIIs). Additionally, HyperCard allows us to fully utilise the interactive design/test/refine process in a highly efficient manner. System development proceeded on two fronts: hardware and software capabilities, and issues of intellectual access.

Hardware and software issues:

The system was developed on a Macintosh IIcx and Quadra 700, both running System 7.1, QuickTime 1.5 and HyperCard 2.1. We scanned each image using an 8 bit colour flatbed scanner (LaCie SilverScanner) and its associated software (we choose not to pursue enhancing the quality of our images through the use of image manipulation software at this time). Scanned image files were compressed by a QuickTime XCMD in HyperCard using the JPEG codec and a quality setting of 128. This typically reduced a 60KB file to 12KB with no detectable loss of image quality. We are experimenting with digitising video footage of figurines to provide a 360 degree record. The HyperCard stack which comprises the system is capable of displaying the image files in independent, scrolling document windows directly from disk. There are approximately 200 images in the current database. The 8 bit gray scale images are stored as separate PICT files.

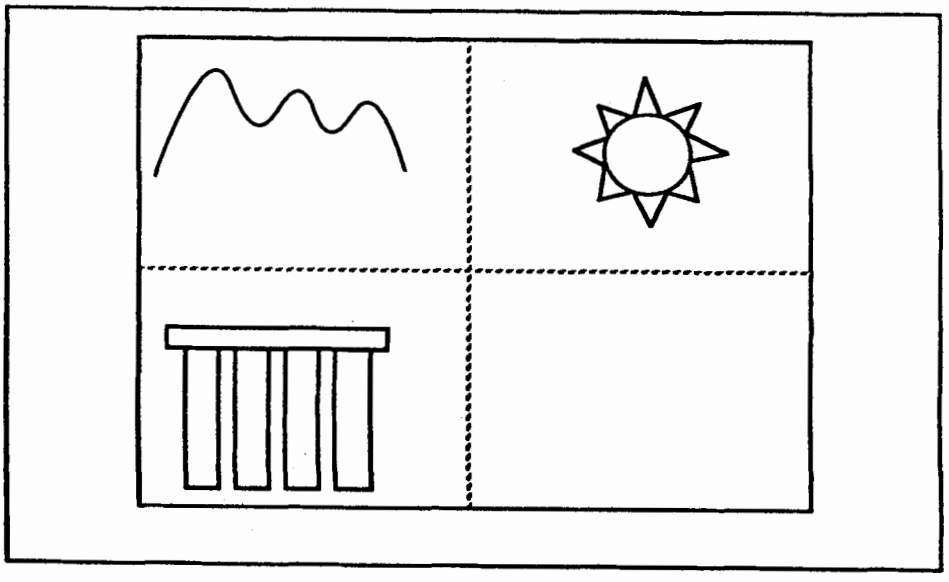
Textual description is in English or Italian in standard fields: catalogue number, plate number, the inventory number, condition, fabrication techniques, measurements, series,

Fig. 1 Typical text description

The Visual Thesaurus

Head of Athena figurine. Wears helmet with high lophos, raised cheek pieces, and two rosette discs at edges over a low stephane. Rosette earrings. Centrally parted hair falls in crinkly waves behind ears and over shoulders. Two sections of locks pulled up over centrally parted strands at sides of head. Blocky face with low forehead, long nose, fleshy lips, and full jaw. Two Venus rings on neck.

Fig. 2 An image broken into quadrants



description, excavation, publication, comparisons, date, etc. The language is chosen from the "Language" menu. A typical textual description is shown in Fig.1. Each card can be linked to one or more image files and any image can be displayed at any time allowing quick retrieval for comparisons.

Intellectual access issues:

Much current research in image retrieval concentrates on providing the searcher with the ability to retrieve large numbers of images directly and then use browsing tools to choose which images are appropriate. This bears a similarity to trends in textual retrieval which rely primarily on keyword searching techniques and ignore older and more conventional approaches such as human indexing, thesauri, and other measures of authority control. We are all familiar with the many problems involving both precision and recall which occur in bibliographic retrieval systems, and we would propose that similar problems will occur in image retrieval systems which rely on a similar metaphor to "keyword" searching. Interestingly, there seems to be little attention being paid to indexing of images (Cawkell, 1991), and textual thesauri such as the Art and Architecture Thesaurus remain time-intensive to apply.

We propose two new approaches to indexing of images, "iconic indexing" and "spatial indexing". Iconic indexing is based on the old adage "a picture is worth a thousand words", and is designed primarily as an aid to the initial cataloguing of a large number of images where a manageable number of features can be successfully represented by an icon. For our collection of terra cotta figurines, these features include human/animal, various positions (seated, standing), body parts (head, arm, foot), and aspects of "dress" (helmet, wigs). In a more general art collection, such categories as portrait, landscape, abstract are, or still life could be iconic index terms. Each of these icons is linked to a textual description. The primary purpose of iconic indexing is to enable rapid categorisation of a large number of images. The use of icons also provides a method of authority control for non-textual search terms and serves as a thesaurus form which to choose search terms. Spatial indexing assigns textual content identifiers to regions of images. Content includes different facets such as object, colour, or atmosphere. What is unique about spatial indexing is that each term is also assigned a simple spatial identifier (see also Bordogna et al., 1990). Four quadrants are sufficient for many images. Thus a picture of a landscape may contain a tree, lake, and setting sun. What is unique to a particular image is the spatial relationship of these objects to each other (Fig.2). Thus, by assigning a few textual index terms and their quadrant position, a unique record is created which can be searched easily. In a graphical user interface, a user places terms in the appropriate quadrant to form a search statement.

For the purposes of the particular visual thesaurus we are working with, the image will be divided into three segments (roughly corresponding to head, torso and legs); this is the most appropriate division given the positioning of the objects for photographic record.

Textual and Image search features:

The system developers have observed a widespread problem in current computerised retrieval systems: a limited amount of flexibility in the type of searching strategies allowed at any one point in the search. Drawing on results from another research project (Liddy and Jörgensen, 1993), flexibility in retrieval strategies is emphasised in the current project. Search statements can be composed of either text, iconic index images, or both, with Boolean combinations also possible. Search statements are composed by dragging terms into a search box from an alphabetic index list, typing terms in directly, or dragging iconic index (terms) into the search box. Thus the researcher can retrieve all standing figures (iconic index term) from Mold series 1.a.3 (text term).

Searching can also be accomplished using spatial indexing, with either textual terms or iconic index terms placed in the appropriate image segment. This type of flexibility in composition of search statements allows the researcher to retrieve images which have important but perhaps unperceived similarities, allowing for an exploratory approach to data analysis. At the same time, retrieval can be limited for more precise searchers, or to control the number of images retrieved.

Testing in the field

In addition to general observations of system usage, we will be testing different retrieval conditions: text-based keyword searching; text and icon-based retrieval; and text and icon-based retrieval with the addition of spatial indexing. With this research we hope to emphasise the complementary aspect of visual retrieval to textual retrieval, and to define specific functionalities for visual as opposed to textual retrieval.