CONTENT BASED RETRIEVAL AND NAVIGATION WITH IMAGES IN THE MICROCOSM MODEL

Paul H. Lewis, Hugh C. Davis, Steve R. Griffiths, Wendy Hall, Rob J. Wilkins
The Multimedia Group
Department of Electronics and Computer Science
University of Southampton, England, SO171BJ.
E-mail phl@ecs.soton.ac.uk

ABSTRACT

This paper presents an overview of novel extensions to the Microcosm open hypermedia architecture which facilitate content based retrieval and navigation with non-textual media such as images, video and sound. MAVIS (Microcosm Architecture for Video, Image and Sound) extends the area of application of *generic* and *computed links* to non-textual media. Examples of the use of shape and colour for content based retrieval and navigation are presented.

(*This paper appeared in the proceedings of The International Conference on Multimedia Communications, MEDIACOMM'95, pp 86–90, published by the Society for Computer Simulation International*)

INTRODUCTION

One of the major challenges for multimedia information systems (MMIS) developers is the provision of effective tools for navigating through, or retrieving images from, image collections or other non-textual information such as video and sound sequences. Such tools lag a long way behind tools for handling text. Confronted with a multimedia information system, with an image on the screen, we would like to be able to select the image or a sub-image and ask for other images or sub-images which are similar in some way to our particular selection. If we can use the image content of the selection to match in some way with images or sub-images in the MMIS, this is referred to as content based retrieval (CBR). Similarly, asking for other information relating to the image selection and navigating to that information using the content of the selection as the key, is referred to as content based navigation (CBN).

As an example, consider a user of an on-line multimedia furniture catalogue who finds a picture of a pine table in the catalogue and wishes to know if there are other items of furniture which might match the table in that or other manufacturers' catalogues. The user asks the computer to locate any similar images. The system responds with all the other pieces of pine furniture of which it has knowledge. This process is CBR. The user then asks for more information about a specific item such as the price, the delivery date or a video of a house interior with the furniture in place. This process is CBN.

Retrieval and navigation with images in MMIS has been tackled for many years by the use of associated and manually assigned keys or the use of manually authored point to point links. By contrast, techniques which capitalise on the direct use of image content as an aid to navigation or retrieval are in their infancy. Research which can contribute to progress in this area comes not only from the multimedia community, but also from image database research and research into image understanding and computer vision

Published work includes descriptions of image databases which use a range of image properties for direct content based retrieval. Typically they are for use in specific application domains e.g. [2, 3]. Some systems provide content based image retrieval tools of wider applicability but are often less powerful than those tuned to a specific application domain e.g. the QBIC system [6] and the Manchester Content Addressable Image Database [7].

Although these, and other papers address the problem of image retrieval by content either with or without semantic considerations, few papers have explicitly addressed the problem of content based retrieval within an MMIS environment and fewer have addressed the general issue of content based navigation through an MMIS. A notable exception is the Miyabi system [5], under development at NEC, which is a hypermedia system in which non-textual information may be retrieved using queries presented in the same form as the medium to be retrieved.

In spite of significant progress in content based retrieval, it is clear that to achieve truly robust and versatile retrieval of images based on content will require a general solution to the recognition problem in image analysis and this is far from being achieved. However, parameterisations of shape, colour, texture and other image features can provide useful content cues for direct addressing of images and provide valuable filters for image matching and retrieval.

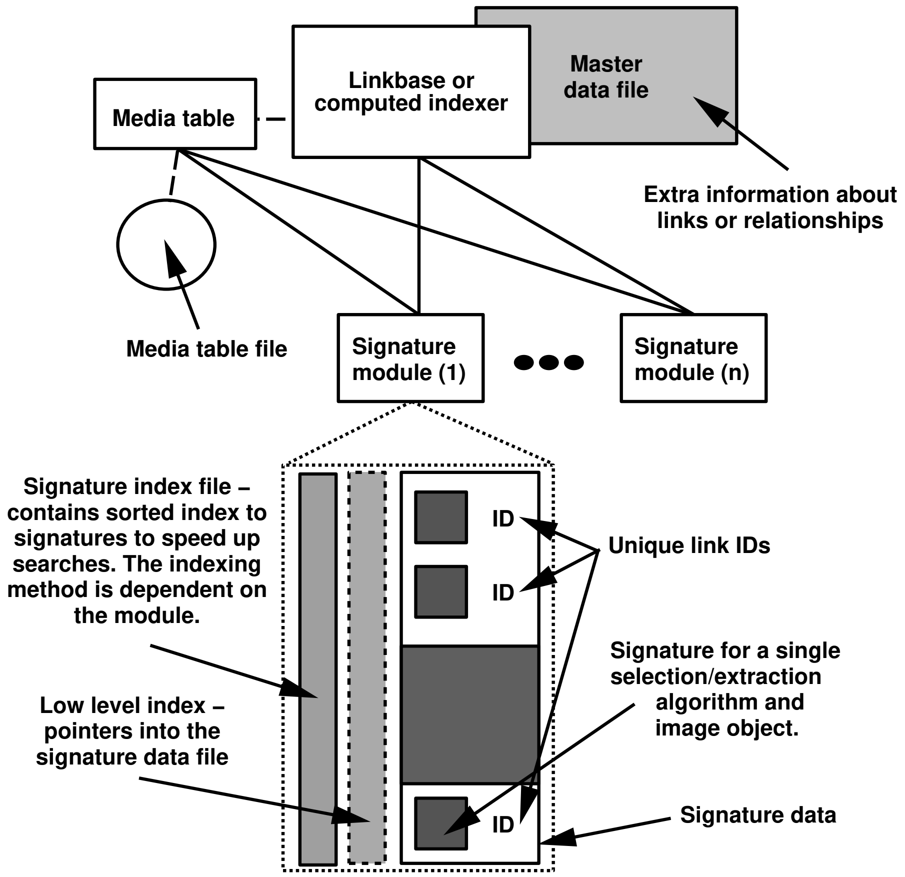

Figure 1: Module Framework for CBR and CBN

In the following sections we briefly describe the Microcosm open hypermedia architecture and our recent extension to support content based retrieval and navigation.

THE MICROCOSM MODEL

Over the last five years, the Multimedia Research Group at Southampton University has developed Microcosm [1, 4], an open architecture MMIS, which is not only a valuable vehicle for MMIS research but also a versatile open hypermedia system in which all datafiles remain in the native format of the application that created them. All link information is held in separate link databases (linkbases).

Link authoring within hypermedia applications is an arduous and time consuming process. As the size and complexity of the hypermedia applications increase the number of links needed to relate the disparate sections of information (or documents) also increases. Authoring every link by hand soon ceases to be feasible and Microcosm provides two powerful alternatives to the manual creation of every link. These are:

- **Generic links** - A generic link may be followed from a given object at any location in any document. For example, one generic link could be authored from the text string *Southampton* to a video of the city. Wherever this string occurs in any document, the user can then follow the link to the video.
- **Compute Links** - Text documents within the application may be pre-indexed to allow full text retrieval searches. This is integrated into the link following environment so that the user can then select any piece of text and ask the system to attempt to provide related information according to a predefined algorithm. The facility is called *compute links* as it allows users to navigate through the information without following predefined links.

There is no doubt that both of the above mechanisms reduce the authoring load associated with the creation of large complex hypermedia applications. The former provides a method for replacing multiple links to the same destination with a single generic link and the latter provides a tool for relating documents without needing any links to be authored. However, until recently, these two mechanisms for reducing authoring load were only available for text. In the next section we provide an overview of recent extensions to the Microcosm architecture producing MAVIS (Microcosm Architecture for Video, Image and Sound) which

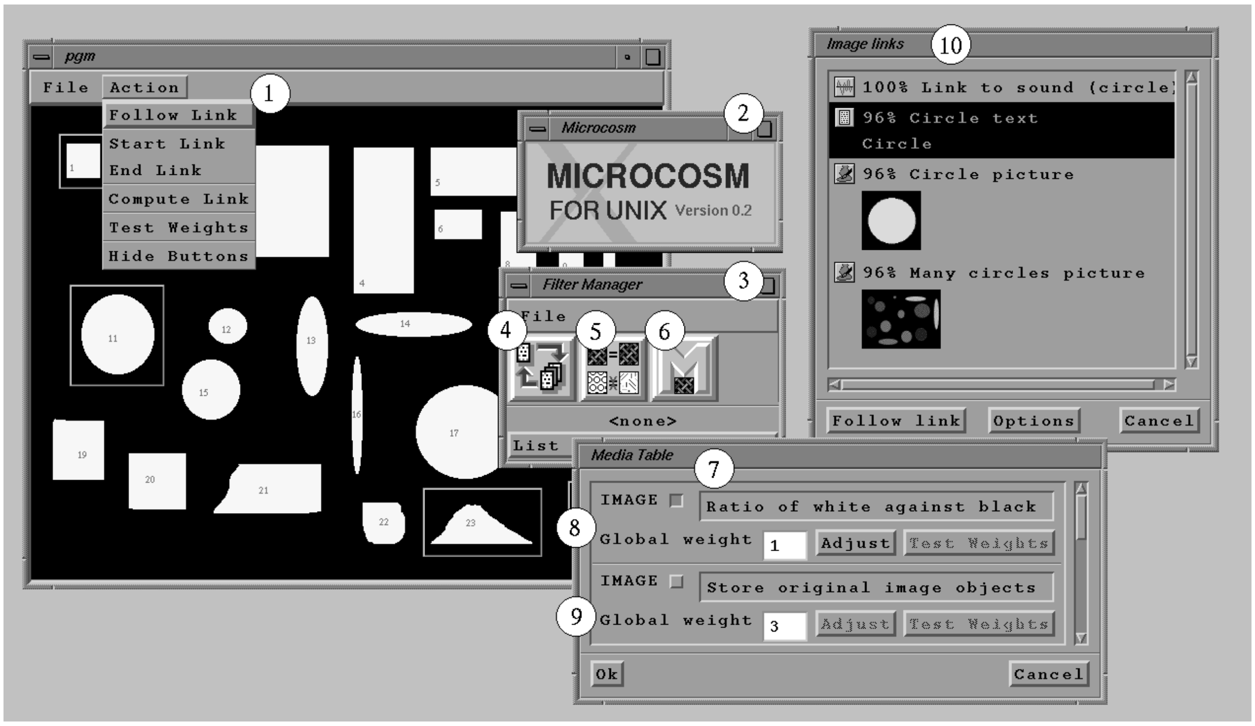

Figure 2: Screen Shot Showing Content Based Navigation

allows authors to exploit the mechanisms for *generic links* and *compute links* on non-textual media.

EXTENSIONS TO THE ARCHITECTURE

In the standard version of Microcosm, links are authored by selecting areas of documents to act as start and end anchors of the link and then choosing a create link action. The linking process within Microcosm then allows the author to select the type of link that they wish to create such as a generic link or button. Links from text documents contain the selection that constitutes the link's start anchor. However, links from non-textual media do not. For example, a link from an image will contain the bounding box describing the selected area but not its content. Although this allows specific and local links from images, the lack of content rendered it difficult to implement generic links from non-textual material within the existing Microcosm architecture.

In MAVIS the architecture has been extended to allow links from non-textual media to include the content of their source anchors. This elegant extension gives the system the potential to exploit generic links in non-textual documents.

The extension takes the form of a framework containing a collection of modules. Unlike text, where the only comparison used to retrieve links is a straight forward string match, richer media such as images require many different matching algorithms. Each module in the framework is responsible for maintaining everything needed to produce and maintain indexed representations (called signatures) of the content from the link start anchors for one matching process. For example, images can be matched using such features as colour distribution, shape and texture. In our system, each feature is handled by a separate module in the framework.

Each signature module maintains the signature data (processed representations of the link start anchor content) and a signature type dependent index. They also include an algorithm for matching signatures and generating similarity indices. The MAVIS architecture is shown in figure 1.

Navigation (CBN)

In the new system, when an author creates a link from an image, the image viewer passes the selected area to the linkbase. The linkbase then passes this to each active module and they process the selection, adding the resulting signatures to their indexes. The interface for the novice user, for creating links, is identical in the standard and the extended systems.

When a link is followed, the image viewer passes the selected area to the linkbase once more. The linkbase passes the selection to each active module which processes the selection to produce a signature and then matches this against all similar signatures in its index. Each module returns a list of the best matches to the linkbase. The linkbase merges these lists and then continues in the same way as the standard system.

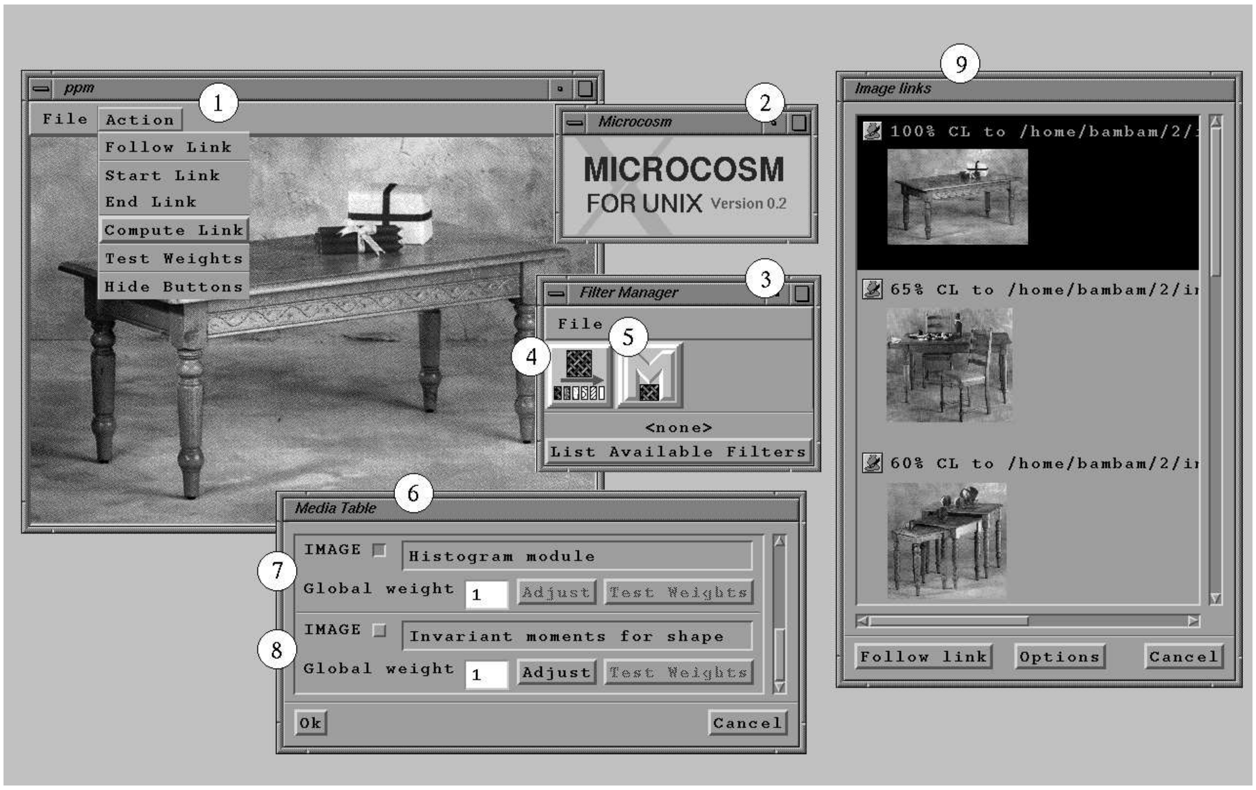

Figure 3: Screen Shot Showing Content Based Retrieval

Retrieval (CBR)

Content based retrieval is the equivalent of full text retrieval but for non-textual documents. Pre-indexing of the document using specific aspects of its content takes place. For example, an image may be indexed on colour distribution or object shape, whereas a sound document may be indexed using parameterisations of its constituent frequencies.

The MAVIS extensions include a framework for CBR from non-textual documents, in addition to the generic link facilities described above. The facility uses the same modules as those for generic link operation but without the need for manual creation of relations between regions of the documents. Instead, feature selections from the documents are used to generate the link entries, and any matches found will return the document (image, video or sound clip) corresponding to the match.

EXAMPLES OF NAVIGATION AND RETRIEVAL

In this section we present two illustrative examples of MAVIS in action. In the first, a *generic link* is being followed to find further information about the selected object, demonstrating the concept of content based navigation.

In the second, the *compute link* option is used to retrieve images similar to the one selected, demonstrating content based retrieval of images.

Follow Link (CBN)

Figure 2 shows a user following a link from a selection in an image containing simple black and white geometric shapes. The user has selected one of the circles under the menu (1) and then chosen the *Follow link* action from the menu. The content of the selection is passed to the MAVIS linkbase (5). The linkbase passes on the selection to any active modules (as shown in (7)). In this example, the user has chosen to make module (8) active and module (9) inactive. Module (8) generates a signature based on the percentage of white in the selection and matches it against signatures in its index. The results are presented in the *Image links* window (10). It should be stressed that the results show authored relations between an object that is similar to a circle (i.e. any circular object) and the items displayed in window (10). This type of link can be followed from any object that matches (i.e. is similar to) the start point of the link (in this case the circle). However, the link can be to any media type (e.g. sound, video, text or images).

The other numbered items in the figure are the main Microcosm process(2), the filter management part of Microcosm(3), a filter that allows the user to create links between objects(4) and a filter that shows the user available links to follow from their chosen object(6).

Compute Link (CBR)

Figure 3 shows a user searching for objects that are similar to a selected object (in this case the object is the pine table under menu (1)). The user has selected the whole image to be treated as the query object. They have then chosen the *Compute link* option on menu (1). The content of their selection is passed to the MAVIS CBR filter (4). This filter then passes the selection to all of its active modules (as shown in (6)). In this example the user only chose module (7) as active. Module (7) generates a signature based on the colour content of the selection and matches it against other signatures in its index. The results are presented in the *Image links* window (9). It should be stressed that the results show objects that are similar to the user's selected object. No links have been authored, all that is needed is to pre-index the images for each module (in this case only the colour content module).

The other numbered items in the figure are the main Microcosm process(2), the filter management part of Microcosm(3), a filter that shows the user available links to follow from their chosen object(5) and an inactive module(8).

NAVIGATING FROM OTHER MEDIA

The extensions can, in principle, handle matching processes from any media types. The new architecture contains a *Media table* which relates media types to signature modules. This allows new media types and signature modules to be added dynamically to the system. If a particular set of objects require a specific matching algorithm then the relevant signature module can be constructed and plugged into the system. In addition, if a new media type is to be integrated then the *Media table* can be dynamically extended to associate it to existing or new signature modules.

CONCLUSIONS AND FUTURE WORK

In this paper we have provided and overview of MAVIS, a novel enhancement to the MICROCOSM architecture, which extends the applicability of *generic* and *compute links* to non-textual media and provides an elegant framework for content based retrieval and navigation. Exemplars of content based retrieval and navigation for images have also been presented.

Future extensions will allow the use of arbitrarily shaped selections both for authoring links, and for use by the computed indexer. The latter will require either manual selection of the features from each image to be used, or an automated (or semi-automated) segmentation routine to obtain features suitable for indexing. Arbitrary shape selections can already be handled to some extent within the current data structures used to process images, but these features are not yet included within the modules.

The power of the content based retrieval and navigation facilities is ultimately determined by the extraction and matching algorithms within the feature modules. We are currently working on a variety of additional modules to increase the range of tools available within the architecture.

ACKNOWLEDGEMENTS

This research is being undertaken with the support of SERC grant number GR/J53614.

References

[1] H. Davis, W. Hall, I. Heath, G. Hill, and R. Wilkins. Towards an integrated information environment with open hypermedia systems. In D. Lucarella, J. Nanard, Nanard M., and P. Paolini, editors, *Fourth ACM Conference on Hypertext (ECHT '92)*, pages 181–190, Milan, Italy, December 1992.

[2] T. Gevers and A.W.M. Smeulders. Σnigma: An image retrieval system. In *11th Internationl Conference on Pattern Recognition B*, pages 697–700, The Hague, Netherlands, 1992.

[3] A. Gupta, T. Weymouth, and R. Jain. Semantic queries with pictures: The VIMSYS model. In *Proc. 17th International Conf. on Very Large Data Bases*, pages 69–79, Barcelona, September 1991.

[4] G. J. Hill, R. J. Wilkins, and W. Hall. Open and reconfigurable hypermedia systems: A filter-based model. *Hypermedia*, 5(2):103–118, 1993.

[5] K. Hirata, Y. Hara, N. Shibata, and F. Hirabayashi. Media-based navigation for hypermedia systems. In *Hypertext '93*, pages 159–173, Seattle, Washington, USA, November 14-18 1993. ACM Press.

[6] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, and C. Faloutsos. The QBIC project: Querying images by content using colour, texture, and shape. Research Report RJ 9203 (81511), IBM Research Division, Almanden Research Center, San Jose, California, February 1993.

[7] J. Oakley, R. Shann, D. Davis, and L. Hugueville. A database management system for vision applications. In *5th British Machine Vision Conference*, pages 629–639, York, U.K., 13-16 September 1994. BMVA Press.