Intelligent user interface for an content-based information retrieval system in schools

OLAF NEUMANN / ROBERT HESS / ILONA KUNZ
Department of computer science
Dresden University of Technology
Hans-Grundig-Str. 25 - 01062 Dresden
GERMANY

Abstract: - A multimedia information retrieval system goes beyond traditional database systems to incorporate various modes of non-textual digital data, such as digitized images and videos, in addition to textual information.
In Modulates it is intended to search for images or videos to access additional material in the multimedia courses provided. A major goal is the integration of the content based information retrieval user interface in the learning system.
This paper gives an exhausting overview over existing systems for content based information retrieval concerning the capabilities of their user interfaces.
Based on the discussion of these interfaces a proposal for a new interface between a learning system and an arbitrary content based retrieval system is derived. Special considerations related to school learning aspects as the concept of user profiles for different degrees of expertise are introduced.
The paper delivers information about how a state of the art technology can be used in the environment of schools in order to enrich the computer assisted learning process. Advantages and disadvantages for several application fields are discussed.
Key-Words: content based retrieval of visual data, image data bases, educational multimedia systems
CSCC'99 Proceedings:- Pages 6371-6377

1 Introduction

A multimedia information retrieval system goes beyond traditional database systems to incorporate various modes of non-textual digital data, such as digitised images and videos, in addition to textual information.
In the EU founded MODULATES[1] project, it is intended to search for images or videos to access additional material in common multimedia courses provided by the project partners. A major goal is the integration of the content based information retrieval user interface in the learning system.
This paper gives an exhausting overview over existing systems for content based information retrieval concerning the capabilities of their user interfaces.
In the Scope of content based visual data retrieval in the field of educational applications, the following topics are investigated in detail within the MODULATES project:

- image (sketch) recognition,
- image based queries (using image databases),
- queries over video databases,
- user interfaces for the above mentioned techniques.

Focus of this paper will be the question of an appropriate interface for the school domain, since we find very specific requirements here.
It is presumed, that the reader has some basic understanding of techniques and terms of content based visual information retrieval. Otherwise refer to the excellent description in [1], please.

2 Scenario

The following Scenario for content based information retrieval will be used within MODULATES. Within a multimedia course the option will be given to search for similar images for a given image in the course material by simply selecting the image and choosing a menu entry ”search for similar pictures”, e. g. in a motor car course we will search for similar ”red” cars.
The image and some additional parameters are passed from the course's runtime system (in this case authorware) to the query construction module. With these parameters information about the current user can be fetched from the controlling user management system which keeps a so called user profile.
Based on preferences from the user profile and parameters passed by the multimedia course the fitting features are selected and weighted accordingly. Then the query is issued over the network to the query engine. In return a

[1] MODULATES stands for *Multimedia Organisation for Developing the Understanding and Learning of Advanced Technology in European Schools* and is funded by the *EC Educational Multimedia Task Force under Contract MM1018-MODULATES*

list of matching images is passed. It is also possible, that these matching images are key frames from video scenes, that contain appropriate material, e. g. cars in a motor car racing.
The user profile also contains information on whether the user is at beginner or advanced level. Based on this, the interface presented to the user in the following steps is different. Beginner only see a selection of images that are returned by the query and than click on the nearest match(es) in order to refine their search. It might be even possible to go on without refining, when the parameters passed by authorware were detailed enough for a good first matching.
Being an advanced user, there is a multitude of parameters that can be used to refine the search for better matching pictures. Here the possibility to use keywords, provide sketches and insert external images as a starting point for the search is given. Here it is also possible to select a specific domain for the search. Thus the concept of content based retrieval can be explained and used in several degrees of complexity.
Of course there is some processing involved after finding one or several fitting images. The user is usually presented only a number of small thumbnails, representing larger images or even video scenes. After finding some fitting items, the user will probably want to see the images in original resolution, or the entire video clip. Perhaps he also wants to see other scenes from the video. And of course the user wants to store these search results, either as an address to find them again later or even the entire image / clip for further processing.
This scenario is, of course, only one way to exploit the capabilities of content based information retrieval. Other ideas derived within the project can be found in [2].

3 State of the Art

There are several problems to solve. Some of these problems are:

- To query in an appropriate way (interface)
- To search for the queried items in whatever data structures there might exist
- To represent the query results, especially when these are numerous
- To build the above mentioned data structures from whatever kind of content there might be (images, video, animations, text)

Besides these problems, there is a multitude of others. For instance, performance problems, such as computing time and network bandwidth required, or problems from the application domain, starting from legacy content and ending up with pedagogical problems when used in schools, e. g., when is such an interface really more useful than the much cheaper text based, or what someone is looking for in a video.

Up to now, there is a wide range of proposed algorithms to tackle several of the four problems mentioned above. But on the other hand, this research is too new to provide extensive field testing of these algorithms. Most parts of the MPEG 4 standard have entered the final stage, while others, such as conformance testing are just reaching it. Thus the first step to do is an extensive survey regarding the state of the art.

3.1 Existing User Interfaces

The interfaces of a whole range of existing systems, both experimental and commercial, were examined. The results of these examinations are used later in this paper to design a new interface, that is both, expressive and powerful in the construction of content based queries, but also easy enough to be used by pupils.

3.1.1 Image

The retrieval of images according to their content is one of the most mature fields in content based information retrieval. That means, there are several experimental systems available and even a few systems with product status. Nearly 20 applications realised with these systems were examined with respect to their user interface. The systems were chosen to give an comprehensive overview over the current state of the art. Screen dumps and detailed descriptions for every system can be found in [2]. A classification was used to rank the user interfaces. This is of, course, a very subjective view. Since most of the systems are far from optimal, it was decided to create a dedicated interface for MODULATES, combining several useful features of other interfaces.

3.1.2 Classification

Today's system for content based image retrieval offer a wide facet of user interfaces. At closer examination these systems could be classified as follows:

- Systems allowing for the query accordingly to a single feature (i.e. colour representation, or texture)
- Systems allowing to combine several visual features
- Systems trying to exploit spatial features in addition to colour and texture
- Systems that use well defined annotations alongside with the above mentioned features
- Systems, that in addition try to exploit other contextual information not directly contained in the object, but adjacent to it (Audio tracks in video, text on the web site containing an image)

However, these criteria reflect partly the functionality of the underlying retrieval engine, so that an interface could not be estimated based only on these abilities. Thus the following systems were ranked (see Table 1 & 2) according to the following criteria:

- *Ease of use:* Self-explaining functions, clearly structured elements, use of common interface elements

in the usual way.

- *Expressiveness of the query:* How complex are the search operations the interface allows, for instance editors for features, number of supported features, sophisticated combination of features.
- *Responsiveness:* (as far as this can be evaluated given the nature of the internet) Caching of retrieved pictures, local operations for editors, time to calculate features for new pictures
- *Relevance Feedback:* What means of relevance feedback are integrated
- *Possibilities to store / recall search results:* Clipbooks, support for copy or save
- *Ways to search for external images:* Support for load, paste or drag & drop
- *Ways to extend the search database:* Insert images or add URL's

System	Overall rating	Ease of use	Expres-siveness	store / recall search results	Respon-siveness	Relevance Feedback	search external images	extend the search database
Altavista	-	++	--	n.a.		-	n.a.	n.a.
C-Bird (old interface)	-	-	+	n.a.	--	-	URL	n.a.
C-Bird (new interface)	++	+	++	n.a.	+	+	n.a.	n.a.
CIIR	++	+	+	n.a.	+	+	n.a.	n.a.
French Institut for Culture	-	+	+	n.a.	-	-	n.a.	n.a.
Image Rover	+	++	+	+	++	-	n.a.	n.a.
Photobook	++	++	-	n.a.	-	+	n.a.	n.a.
Photo-Demo	+	+	+	n.a.	-	+	n.a.	n.a.
QBIC-HTML-Interface	+	+	+	n.a.	+	+	URL	n.a.
Surfimage	+	+	+	n.a.	++	+	n.a.	n.a.
Virage	+	+	+	n.a.	+	+	n.a.	n.a.
Visinfo	+	++	+	n.a.	++	+	n.a.	n.a.
WebSEEk	+	++	+	n.a.	-	++	n.a.	n.a.
Excalibur Image Surfer	-	+	+	n.a.	+	-	n.a.	n.a.
MetaSEEk	+	+	+	n.a.	+	+	URL	n.a.
Content-based Image Retrieval System	-	+	+	n.a.	-	+	n.a.	n.a.
SaFe	-	--	-	n.a.	-	+	n.a.	n.a.

*Table 1: Rating of interfaces for content based information retrieval ++ very good, - - very bad, **n.a.** not available*

System	URL	Engine	Interface
Altavista	http://image.altavista.com	Virage	
C-Bird (old interface)	http://jupiter.cs.sfu.ca/cbird/CBIRD.cgi		html QBIC-like
C-Bird (new interface)	http://jupiter.cs.sfu.ca/cbird/java/cbirdonline.htm		Java,
CIIR	http://cowarie.cs.umass.edu/~demo/Demo.html	SYNAPSE	
French Institut for Culture	http://www.culture.fr:8099/cgi-bin/QbicStable	QBIC	
Image Rover	http://pacific.bu.edu:7501/cgi-bin/irquery	(cgi)	
Photobook	http://www-white.media.mit.edu/vismod/demos/photobook/	Vismod	
Photo-Demo	http://wwwqbic.almaden.ibm.com/cgi-bin/photo-demo	QBIC - IBM	
QBIC – HTML-Interface	http://wwwqbic.almaden.ibm.com	QBIC - IBM	
Surfimage	http://www-syntim.inria.fr/htbin/syntim/surfimage/surfimage.cgi	IMEDIA Project at INRIA	
Virage	http://www.virage.com/virdemo.html	Virage	
Visinfo	http://visinfo.zib.de/IRS	QBIC	
Webseek	http://www.ctr.columbia.edu/webseek/	Webseek	
Excalibur Image Surfer	http://isurf.interpix.com/	Excalibur Technolo-gies Corporation	
MetaSEEK	http://mahler.ctr.columbia.edu:8080/cgi-bin/MetaSEEk_cate	Virage	
Content-based Image Retrieval System	http://maya.ctr.columbia.edu:8088/cbvq/	J. R. Smith	
SaFe	http://disney.ctr.columbia.edu/safe/	J. R. Smith	

WebClip	http://www.ctr.columbia.edu/webclip/		
CANDID	http://www.c3.lanl.gov/%7Ekelly/CANDID/main.shtml	CANDID	

Table 2: The Systems with their Locations and underlying Engines

3.1.3 Conclusions: Complexity

It could be expected, that the complexity of the user interface grows with the complexity of the possible query. Nevertheless we can see first approaches to encapsulate the complexity of a content based query with several features within a very simple interface (e.g. AltaVista Photo Finder). Unfortunately these encapsulation lead, as a rule, to a significant loss in result quality.

Nevertheless it has to be stated, that using a complex interface may not necessary lead to good results, since there is a variety of parameters, that has to be adjusted not only accordingly to the nature of the sample image, but also accordingly to all the other (unknown) images in the system; this can be only done by guessing or, at best, by experience. This is an even more serious issue, when we consider users, that have no clear understanding about the concepts of content based information retrieval, as pupils usually do, at least in the beginning.

3.1.4 Conclusions: Specialised vs. General Purpose Systems

Another point that should be mentioned is the nature of the images used by these retrieval systems. There are two categories: specialised systems covering only on a specific area (VirInfo, Trademark Database) and general purpose systems as for instance web search engines (AV Photo Finder, WeebSeek) that search over whatever content there may be.

Naturally the chance to find matching images is higher in the specialised systems, since here a pre-selection has been applied and the choice of similarity criteria was made accordingly to the content that could be expected.

On the other hand, the general purpose systems would be more useful for an explorative behaviour, where it is not the primary concern to find *one* specific image, but rather to see what else might be correlated to a piece of information. This is the kind of behaviour we can see for instance, when an expert user tries to explore a subject using an Internet search engine like Alta Vista. We have to stress again, that this procedure differs in its explorative nature significantly from the search for a specific single item of information.

These differences lead to the conclusion, that accordingly to the nature of the specific learning process, different means for content based information retrieval are needed. When getting familiar with this technology and when dealing with very narrow fields of knowledge, it would be quite useful, to limit the scope of the search to a carefully chosen amount of material. This would lead to a higher rate of successful retrievals and would give a positive feedback about the use of this technology. Consequently pupils could learn, what exactly is the use of such a technology and how to use it. Thus they would be prepared for the next steps, when more and more degrees of freedom in the search of material are offered and eventually a very limited choice of system parameter is possible. The later would assume at least a basic familiarity with content based information retrieval concepts which is not required in the first steps.

The above are issues which will be taken into account in the development of the MODULATES system.

3.1.5 Video

Alta Vista also features a search engine based on the Virage video cataloger containing the Clinton testimony. Here it is possible to enter phrases or terms and the video sequences containing these are represented for playback. Since this is based on a full transcript of the session, no real content based retrieval is involved here.

Another video search engine is VideoQ presented at Columbia University which expands the traditional search methods (e.g., keywords and subject navigation) with a novel search technique that allows users to search video based on a rich set of visual features and spatio-temporal relationships.

Obviously this field is not so well developed as in the case of image retrieval. Thus there are only few additional features besides the ones from image retrieval. The major new achievement is the introduction of movement and the according parts of the interface, as presented in VideoQ.

3.1.6 Summary

From the majority of these examples it seems, that such an interface will not necessarily be simple to use nor will it be sufficient to supply only one-medium input for efficient search. It is on the contrary the multitude of media, that can be combined into one query, and the successive refinement of these queries, that promise an advantage over common text based interfaces. At the same time, this multitude introduces the above mentioned complexity.

The challenge for the MODULATES project is, to design an interface, that is both, complex enough to allow for effective queries and simple enough to be used by pupils.

3.1.7 Future Developments

The field of content based information retrieval shows rapid development and can be compared to the development of the text based retrieval machines as Alta Vista or Yahoo a few years ago. So we can hint here only on to important developments instead of describing the

whole very broad field of emerging new applications and technologies.
Virage, Inc. will use IBM's speech technology to bring audio cataloguing and analysis products to the broadcast industry and corporate media management markets. This is intended to solve partly the problem of how to index the waste mass of media available at broadcast companies, in order to prepare it for retrieval.
Magnifi will use IBM's QBIC software in its products. Magnifi delivers a browser-based enterprise application for automatically locating, indexing, retrieving, and visualising data files and rich media information of all types across a network, including stored and streaming audio and video found in media and entertainment companies. With Microsoft, Magnifi will provide technology to complement NetShow with content based query capabilities.
These developments show clearly, that content based information retrieval will become really widely used within the next years.

4 Design of an optimal system for educational purposes

The results of these examinations are used now to design a new interface, that is both, expressive and powerful in the construction of content based queries, but also easy enough to be used by pupils.

4.1 Requirements

From the previous chapters, the following requirements on a content based information retrieval system for educational purposes are derived:

- User interface should support the specific learning styles, that is searching in narrow domains respectively free exploration of large sets of information (see discussion on: Specialised vs. General Purpose Systems 3.1.4).
- The complexity of the interface should be adequate to the users understanding of content based information retrieval and to their needs to interfere with the behaviour of the underlying search engine.
- Automatic selection of features and their weighting that fit the selected search domains, so that the user has not to bother with details of the retrieval process. Here intelligent agent-concepts will provide the necessary information to customise the query.
- Queries need to be saved to reproduce specific query results. This might be crucial for verification by the teacher or for the preparation of exercises.

4.2 System Architecture

It is necessary to include the facilities for content based information retrieval directly into the courses, in order to facilitate a simple handling of the system. Since authorware does not allow a seamless integration, an auxiliary interface will be used to view the search results and to refine the them. (This interface can also be used to control a standalone search application e. g. for the use by course designers or teachers, that want to explore the content more freely.)
Nevertheless it is crucial for the quality of the results, to use as much context information as possible. So at least the topic of the course from which the query was issued is known, and, depending on the information provided by the course designer, a list of keywords related to the specific image.

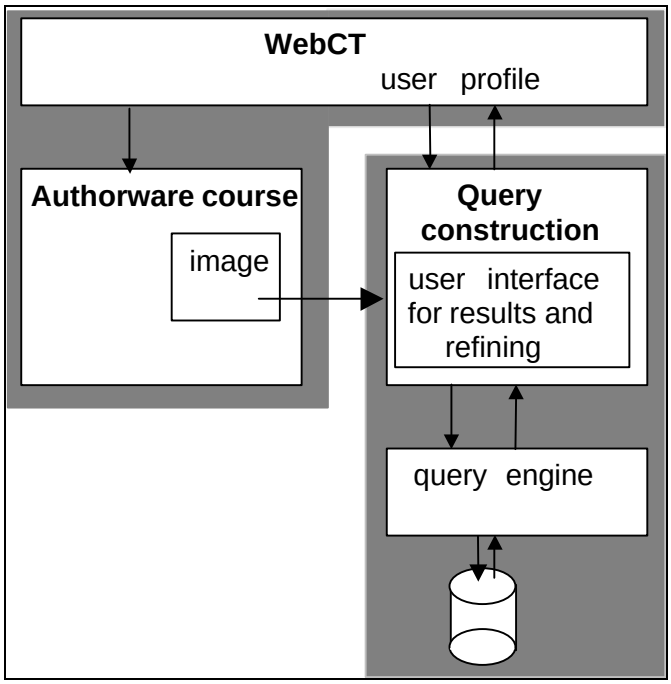

Fig. 1: System Architecture

On the other hand, there is a certain context information, that can be directly derived from the user. At the basic level, this is the information about the users skills (Beginner vs. Advanced) but it is of course also possible to include information about special fields of interests, language preferences and recently found images. Here special care has to be taken, that such preferences do not hinder or confuse the user, by modifying the systems behaviour in an inexplicable way. Research on this field will be carried out in WP2 under the topic of intelligent agent interfaces. The results will be included in the overall system as described above.
The architecture shown in Fig.Fig. 1 defines the main components and interfaces of the system.

4.3 Interface Design

The interfaces for passing information between authorware, WebCT and the query construction module are currently under implementation. The user interface

exists as a first prototype (s. Fig. 2) and the basic ideas can be explained with this.
The right image bar displays the results of the last query. Here images can be selected for viewing or for modification in the white area. Basic modification supported will be the definition of a sub area of a given image by drawing rectangular selection. This modified area will be used as new image to issue another query. It is also possible to import external images as starting points for new queries.
The white area can be used also for the drawing of simple sketches, this shall support the planned feature of sketch recognition. It is not intended, to provide a full featured drawing tool here, but rather simple facilities for drawing lines and freehand figures. Complex graphical manipulations on images as might be required by expert users, can be performed by external graphical packages and the images imported into the system.

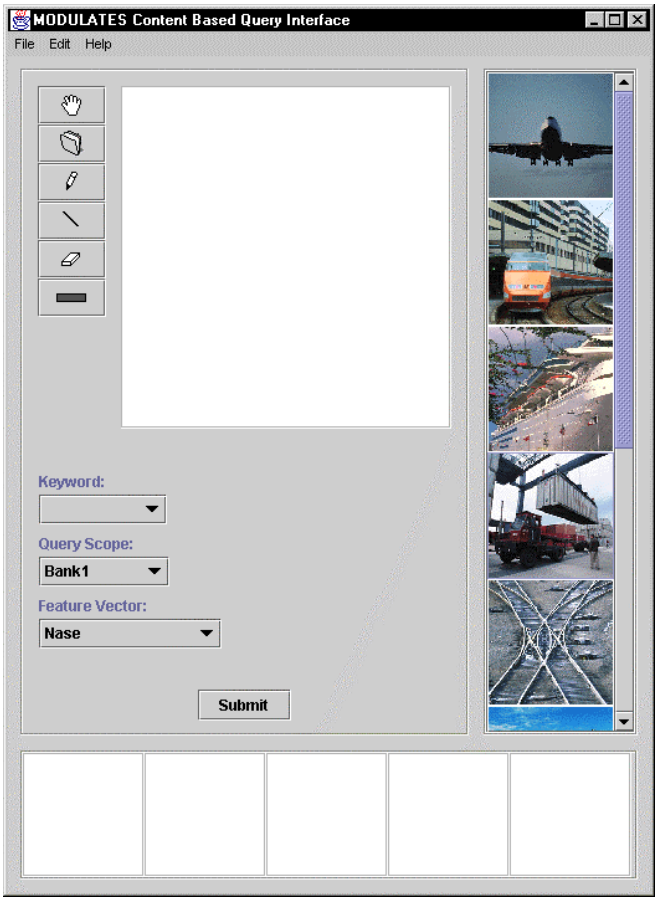

Fig. 2: Prototype user interface for content based information retrieval

Of course, also images from within the database can be exported, manipulated and re-imported as new images.
Furthermore the interface allows the user to select a keyword from the given set of possible keywords, it is possible to narrow the search by specifying a domain, that maps to an underlying classification of the content and, for the advanced user, there might be also the possibility to modify the similarity criterion used for retrieval. The latter option is more for demonstrating the results of using different features, than for real retrieval, where the system should compose automatically a meaningful vector of fitting features and weightings.
On the bottom of the interface, there is a row of free places, where the user can temporarily place search results for later use via drag & drop.

Bibliography and References:

[1] Smith, J.R., Integrated Spatial and Feature Image Systems: Retrieval, Analysis and Compression, PhD-thesis, Columbia University, 1997
[2] Hess et al, Deliverable 3.02 of the MODULATES project.
[3] Niblack et al, ”The QBIC Project: Quering Images by Content Using Color, Texture and Shape”, SPIE Symposium of Electronic Imaging Science and Technology, San Jose, CA, Febr. 1993.
[4] Arman et al, ”Content-based Browsing of Video Sequences”, Proceedings of ACM Multimedia 94, San Francisco, CA, October 1994.
[5] Smoliar and H. Zhang, ”Content-based Video Indexing and Retrieval”, IEEE Multimedia, vol 1, no 2, pp. 62-72, 1994.
[6] Pearson, ”Developments in Model-based Video Coding”, Proceedings of the IEEE, vol.83, no 6, pp. 892-906, June 1995.
[7] L.Chiariglione, ”MPEG-4: Call for Proposals” Document ISO/IEC/JTC1/SC1/SC29/WG11 N820, 1994.
[8] Koenen, F. Pereira and L. Chiariglione, ”MPEG-4, Context and Objectives”, Invited Paper on the Special Issue on MPEG-4, Image Communication Journal, 1997.
[9] F.Pereira and T.Alpert, ”MPEG-4 video subjective test procedures” IEEE Trans.Circuits Syst.Video Technol., pp. 32-51, 1997.
[10] MPEG Integration Group, ”MPEG-4 SNHC Call for Proposals”, Florence MPEG Meeting, Document ISO/IEC/JTC1/SC1/SC29/WG11 N1195, 1996.
[11] L.Chiariglione, ”MPEG-7: Context and Objectives” Document ISO/IEC/JTC1/SC1/SC29/ WG11 N1425, 1996.
[12] Paulo Correia, Fernando Pereira , ”IST video analysis framework” Article from ACcenTS Nr. 3, Vol. 2 - Special Issue on Image Analysis, December 1997
[13] ACcenTS Nr. 3, Vol. 2 - Special Issue on Image Analysis, December 1997
[14] Nack, F., Steinmetz, A., Dittmann J., Wilk, C. ”A4SM, A Framework for Distributed Video Production.” Proceedings of SYBEN98, Zürich, Switzerland, 18-20 May 1998.
[15] Marmolin H., (1990), ”Multimedia from the Perspectives of Psychology”, in Multimedia

Systems, Interactions and Applications, L. Kjelldahl, eds., Springer-Verlag, pp. 39-52
[16] Minka, T. P. ”Master of Engineering Thesis”, 1996. TR #365
[17] Minka T. P. and Picard, R. W. ”Pattern Recognition”, 30(4), 1997. TR #349