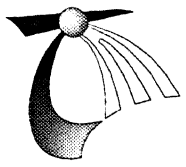

Centre for
Telematics and
Information
Technology

Let's talk about it: Dialogues with Multimedia Databases Database support for human activity

Arjen P. de Vries, Gerrit C. van der Veer and Henk M. Blanken

CTIT Technical Report series, No. 97-13
ISSN 1381 - 3625
P.O. Box 217 - 7500 AE Enschede - The Netherlands
telephone +31-53-4893779 / fax +31-53-4893247

Order-address:

Centre for Telematics and Information Technology
University of Twente
P.O. Box 217
7500 AE Enschede
the Netherlands
email: castaned@ctit.utwente.nl

Let's talk about it: Dialogues with Multimedia Databases

Database support for human activity*

Arjen P. de Vries, Gerrit C. van der Veer and Henk M. Blanken
Centre for Telematics and Information Technology
University of Twente

Abstract: We describe two scenarios of user tasks that need multimedia databases. In these user tasks, access to multimedia data plays a significant role. Current multimedia databases do not provide sufficient functionality for the retrieval of multimedia objects because these objects are stored as passive objects. We discuss several techniques to provide active multimedia objects from a user perspective. We show that querying in multimedia databases should be a dialogue between user and database. We need to combine the information from different media to make retrieval effective. Finally, we introduce a promising database architecture to meet the new user requirements.

Keywords: multimedia, multimedia databases, multimedia modeling, human computer interaction, content-based retrieval, relevance feedback.

1 Introduction

People deal with multimedia data every day. Every time we read a book, watch television or listen to some music, we work with multimedia data. Moreover, we organize and structure this information for ourselves such that we can easily retrieve this information when needed. We create photo albums of our holidays, we possess racks of compact discs and tapes with the music we like, we store past editions of magazines in boxes and use a video recorder to record television programs about topics of our interest. For people with professions like fashion designer or journalist, the amount of information collected is even higher, and the retrieval task is more difficult.

Since the introduction of multimedia in personal computers, we can easily digitize part of our information. This resulted in people creating their own homepages on the world

*This paper was accepted for a full paper presentation at Multimedia Minded 1997

wide web, as a means to manage the information they collect. A major advantage over shoeboxes stored on our attics, is that we can easily share our data collection with others. However, one look at the web makes clear that a computer with a web server is not the best tool to share your shoebox data. It is not easy to find what you want and the information *that* you find is often incorrect or has been moved to another location.

Database technology provides means to store and retrieve high volumes of data. However, until recently, we could not use databases for anything more advanced than names and numbers. Nowadays, we read a lot about multimedia databases. Unfortunately, anything that simply *stores* multimedia data is called a multimedia database. The capabilities of such databases suffice for typical applications of real estate and travel businesses, as these systems only deal with the presentation of otherwise statically used information. A *real* multimedia database is a flexible tool for both storage and retrieval of multimedia data.

2 Example scenarios of user tasks

To illustrate the real-life application of multimedia database systems, we outline two scenarios of user tasks, each demonstrating the functionality that the end user should get from a multimedia database system.

In the first scenario, imagine a journalist writing an article about the effects of alcohol on driving. Before he can start to do the actual work of writing the article, he has to collect news paper articles about recent accidents, scientific reports giving statistics and explanations, photographs, television commercials broadcasted for the government, and interviews with policemen and medical experts.

The second case focuses on a fashion designer developing a concept for a dress to be worn by receptionists of some big retail office. To succeed in this creative design task, he first collects many different multimedia objects. The designer needs descriptions and pictures of the retailer's products, video fragments of buyers at the premises, photographs revealing details of the entrance and reception area, advertisements in magazines, commercials on television, video and audio fragments of sales managers vision development breakfast, and many other pieces of information associated with the retailer. The designer also browses through previous designs, studies preferred dresses from colleagues, and views some videos of recent developments in fashion design.

It is easily understood that the people in both scenarios deal with large amounts of multimedia information to accomplish the tasks they face. Maybe the fashion designer does not need the advanced technology if he works on his own. Piles on his desk and some shoeboxes with old designs may provide easier ways to deal with the data. However, design tasks are typically performed by a team of designers. Even if these people work at the same time in the same room, they would still need a tool to find what they need in the 'organized mess' of the other team members.

3 Searching new media objects

Both user scenarios demonstrate that the key functionality a multimedia database should offer is access to multimedia information. With respect to access, multimedia objects can be classified in two classes: *active* objects and *passive* objects [Bertino et al., 1995]. Active objects really participate in the retrieval process. Users can specify conditions on active objects in the query, referring to the content or referring to the existence. Passive objects just exist in the database. It is not possible to condition on the content of passive objects.

Most information systems that claim to be multimedia databases view images, audio and video objects as passive objects. These databases are not more than huge collections of multimedia data. They do not meet the requirements of the fashion designer or the journalist from the previous section. A tool that just stores data is really not much more than a file system.

In a multimedia database system, all objects should be active objects. We want to use multimedia databases with photo and music collections like we use conventional databases to manage phone numbers. We do not just store the phone numbers and then check all records sequentially each time we want to call John. Instead, we simply ask the database system for John's number. We use a database system as a tool to recollect unknown properties of stored entities using some known properties.

Unfortunately, properties of digitized multimedia objects are not as easily checked as the properties of numbers or strings. Applying an exact match on two digitized objects will only retrieve another object if it is bit-for-bit exactly the same. The question arises why you would search for a digitized object that you already used to formulate the query. It could be useful to find other properties of a multimedia object, similar to searching the phone number using a person's name. Imagine the police officer who needs the name of the criminal he recognized from a photograph. However, in most practical situations we do not have the exact picture that resides in the database. Hence, we need other means to handle the multimedia data as active objects.

3.1 Manually added descriptions

The straightforward approach to using multimedia objects is to manually add a textual description of the object. We know how to search using textual descriptions. An extra advantage of this option is that the search engine would be independent of the media type of the objects in the database.

Obviously, manual indexing is rather expensive if we deal with large amounts of data. However, this approach is problematic in three more fundamental ways. The common cause underlying these problems is the limited capability of capturing the full semantics of multimedia data in textual descriptions.

First, it is not likely that people use keywords to describe objects in a standardized manner. Different people select different words to describe the same concepts. For example, one person may describe a picture of 'an evening in the mountains' as 'dark' while another person describes the same picture as 'somber'. Both try to express approximately the same concept, but if the first searches for the picture in the database collected by the other, he will not find the picture although it is in the database.

We may partly overcome ambiguity in natural language using thesauri. However, the second problem cannot be solved with thesauri. Different people describe different aspects of the picture. The same picture described as 'dark' may be associated with *evening* and *Mount Snowdon* by an enthousiastic hiker. Moreover, even one person will use different descriptions depending on the specific situation when asked. In psychology, this is known as the *encoding specificity principle* [Miller and Johnson-Laird, 1976]. For example, a hiker describes the picture with 'dark' in his office during the week, but he writes down 'evening' in his living room in the weekend.

Finally, substantial evidence exists that many semantic properties of multimedia objects cannot unambiguously be expressed verbally. Several experiments have shown that perception in the right hemisphere is very different from perception in the left hemisphere [Iaccino, 1993]. While the dominating left hemisphere is analytic and verbal, the minor right hemisphere is nonverbal and synthetic.

According to [Barrow, 1995], the composer Carl Orff never admitted a boy to the Vienna Boys' Choir if he already knew how to read and write. Apparently, he believed analytical skills block the creative processes needed to develop musical skills. Similarly, the famous composer Mozart asked his wife to read letters aloud during composing. He was convinced that the analytical part of his mind would be distracted by processing the speech and not disturb the creative part making music. At present, Bettie Edwards developed a new method for teaching creative drawing, based on the insights in the differences between the right and the left part of the brain [Edwards, 1993].

These differences between the hemispheres impose this last problem on the use of manually added descriptions for access to multimedia data. The right brain sees likeliness between things. It senses emotional and aesthetic value of multimedia objects. But the right brain cannot connect these to words. These values may therefore be more easily recognized and compared than described or expressed. The usage of textual descriptions alone to search the database will not retrieve the multimedia objects, because the user uses different valuation processes from the system.

3.2 Approximate retrieval

Another approach to the problem of multimedia search uses automatically derived properties called features [Faloutsos, 1996]. The key to the retrieval process is *similarity* between objects. We search objects that are similar to the query instead of objects that are equal to

Figure 1: Image retrieval based on color features

the query. Therefore, we use the term approximate retrieval as opposed to exact retrieval. Because these features are calculated from the content of the objects, the approach is also known as content-based retrieval.

The features typically describe easy to calculate syntactic properties of the stored objects, as an approach to capture the semantic characteristics that are relevant to the user. The user does not have to know what features the system uses for retrieval. Instead of explicitely dealing with these syntactic features, the user tells the system what kind of objects to search for by giving an example of a good object. We call this query paradigm 'query by example'.

The QBIC (*Query By Image Content*) system [Niblack et al., 1993] introduced this approach to accessing multimedia data in the domain of images. Features used for image retrieval include measures expressing the color distribution of the image. Other features express the texture and the composition of the image. An image query is translated into a point query in some multi-dimensional feature space. The similarity between a query and a database object is estimated using a distance function.

The most common example query to illustrate the approximate retrieval approach uses the picture of a sunset. With this query object, retrieval using color features works very well. However, it is not trivial to find suitable features for the general situation and it is not always easy to judge why the system found that particular object similar. For example, figure 1 demonstrates the retrieved objects if one searches for pictures of red cars. We also retrieve images of buildings or waterfalls, which are semantically completely different. Syntactically though, the picture of a car can be very similar to the picture of a building, if we search in color space alone.

The approximate retrieval approach is not unique for image retrieval. In the Musclefish system, retrieval based on features is applied to the content-based retrieval of generic audio objects [Wold et al., 1996]. Measures based on pitch, energy and more advanced audio properties span feature space. Since the early sixties, a similar approach has been applied to querying full-text retrieval systems in the field of information retrieval [van Rijsbergen, 1979], [Salton, 1989]. Using special purpose speech recognizers, the text retrieval techniques may easily extend to speech documents [de Vries, 1996].

If the features have a clear perceptual interpretation, we may choose to let the user move directly through the feature space. The term navigational querying refers to that situation. Navigational querying has been demonstrated for musicians working with a

database of musical instruments [Eaglestone and Vertegaal, 1994]. Essentially, it is just another way to use the approximate retrieval approach. In the QBIC system, users could directly manipulate the underlying color query. However, it is very hard to find features with a clear semantic interpretation for general multimedia objects and the features are usually not exposed to the user.

3.3 Social information filtering

A slightly different approach to help a user find multimedia objects is described in [Shardanand and M
The underlying idea of this *social information filtering* process is that several people have similar interests. We can collect the judgements of many people about objects in the database, for instance movies or compact discs, and use a nearest neighbour algorithm to find judgement vectors that are similar. Next, the items that appear in a similar vector but have not been judged by the user yet, can be adviced to the user. The technique has been commercialized in the firefly system [firefly, 1996]. After asking you to rate your favorites, firefly gives you a customized list of recommended artists, albums or movies.

Similarity between user judgements has three major benefits over similarity between the objects. First, it overcomes the problems with identifying suitable features for objects like music and art. Second, it has an inherent ability for serendipitous finds. You find objects that you like, but did not explicitly search for. Finally, the approach implicitly deals with qualitative aspects like style which would be hardly possible with automatically derived features.

Technically, it should not be hard to integrate social information filtering with a multimedia database system. To perform approximate retrieval, we already process point queries in multi-dimensional spaces. The difference between both processes comes down to the difference between the space we map objects in and the distance measure among these objects. However, this technique probably only works if the domain for which we collect judgements is rather narrow.

4 New requirements for multimedia databases

In this section, we show that the accessing multimedia data puts new requirements on the database design. In the previous section, we discussed several approaches to handling multimedia objects as active objects. However, these techniques alone are not sufficient to provide multimedia retrieval. We first show that a multimedia database must support iterative search. Next, we discuss the need for a framework to combine the results from different search strategies. Finally, we conclude with the introduction of a promising new architecture suited for multimedia retrieval.

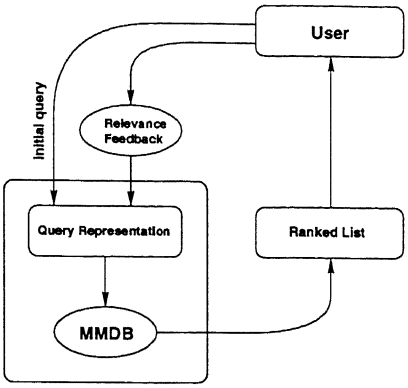

Figure 2: The relevance feedback process

4.1 Interaction with a multimedia database

Interaction with a multimedia database faces us with a major problem that did not exist in the conventional database environment: we do not know how to formulate our multimedia query.

As we already made clear in section 3.1, a multimedia query cannot always be expressed verbally. The query by example paradigm certainly is a major improvement for some retrieval tasks. However, we cannot always come up with an example expressing our information need.

Comparing and evaluating retrieved objects takes place with processes in both hemispheres. Although a user cannot exactly express his information need with a query, the user can judge retrieved objects for relevance. Thus, the solution for the problem of query formulation is to support an iterative search process, see figure 2. After an initial query has been processed, the user is asked to judge the retrieved objects. The relevance judgements are then used to adjust the query to better reflect the user's information need.

Querying multimedia needs a discourse and refinement phase for interaction between the user and the database. Relevance feedback has been used in text retrieval systems [van Rijsbergen, 1979], but not in databases storing arbritary objects. We need to change the database design such that it can use relevance feedback if we want to design multimedia databases.

4.2 Query processing using multiple representations

The techniques in section 3 deal with new media objects like an image or a sound fragment. However, the user is often interested in retrieving composite objects like newspaper articles or video documentaries, so we deal with several atomic objects. Moreover, for each atomic object, we can produce many different representations of multimedia objects. For example, a video fragment can be represented by its subtitles, by the output from

a speech recognizer, or by a sequence of keyframes [Wactlar et al., 1996]. However, we cannot expect the user to search each representation separately and combine the results later.

The usage of multiple representations of multimedia objects is crucial for a multimedia database system. Manually added descriptions are not sufficient for multimedia retrieval. Switching to approximate retrieval techniques overcomes some of the problems with textual descriptions, but introduces new problems because most features have only a syntactic value. Rather than choosing one approach over the other, we should combine the different ways to describe objects in one retrieval engine.

To our best knowledge, no system has used the *multi* aspect in multimedia data. However, multimedia queries cannot be answered by simply checking one aspect like the color distribution of an image. As we saw in figure 1, we retrieve waterfalls and buildings instead of cars. Although a single representation is often not sufficient, the combination of several represenations may be.

4.3 A new database architecture for multimedia retrieval

In the previous sections, we discussed several issues with respect to the access to multimedia data in a database. Summarizing, we list three new requirements for multimedia databases:

- All objects are active objects;
- Querying is an interaction process;
- Query processing uses multiple representations.

Although it is fairly easy to state these requirements, meeting them in an actual system is a tough problem. As an approach to design a system that can meet these requirements, we introduce the architecture of figure 3. We divide the design in a set of agents, a search engine and a conventional database.

Each agent handles one representation of the multimedia objects. For example, one agent produces color features of images. Another agent selects words from the title of text documents. Each agent knows how to find representations that are similar to a the representation of a query object. It creates the necessary access structures in the database to speed up retrieval.

The search engine bridges the database to the user. It keeps track of the different agents that participate in the retrieval process. The subtasks of the retrieval process are delegated to these agents. The search engine contains knowledge about combinating evidence from different representations. The search engine also handles relevance feedback from the user and uses this feedback in further iterations to refine the query.

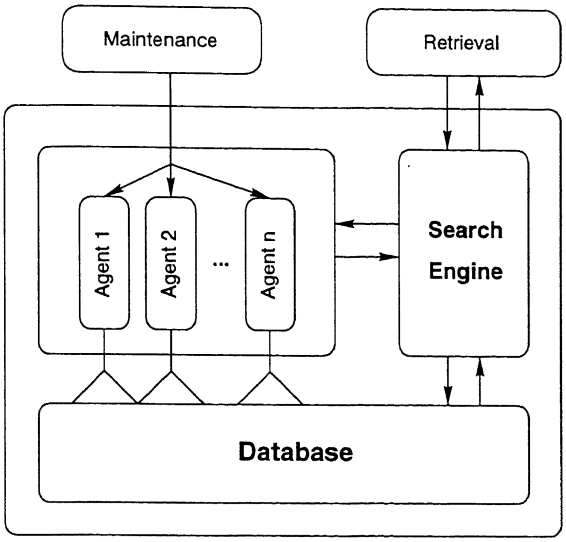

Figure 3: Multimedia database architecture

Agents may process manually added descriptions or features for approximate retrieval. How social information filtering can be used in our architecture is an open question. Theoretically, each agent can store a memory of user judgements for all instances. However, in practice this would be a rather costly solution.

Building the search engine is still a research problem. At present, it is unknown how to combine evidence from different representations of the objects. We need a unifying framework to describe the amount of evidence found for an object by each agent. In probabilistic text retrieval systems, probability theory provides such a framework.

The probabilistic retrieval model ranks the objects by the probability of usefulness to the user given the query. The theory handles relevance feedback through Bayesian probability updating. This probability updating can be performed efficiently if we use Bayesian inference networks [Pearl, 1989]. In the text retrieval system INQUERY [Callan et al., 1992], probabilistic retrieval with inference networks has succesfully combined several textual representations of documents. Our further research focuses on adapting these results for our search engine for multimedia retrieval.

5 Conclusions and further work

A fashion designer and a journalist work with high volumes of multimedia data and they definitely need a flexible storage and retrieval system to cope with their information collections and *especially* with those of their colleagues. In fact, everybody who collects and uses multimedia data is a candidate user of multimedia databases.

In this paper, we identified the properties that a true multimedia database system should

have before we can effectively use computers to replace our bookshelves and shoeboxes. We gave three important requirements on multimedia databases. First, all objects should be handled as active objects. Next, retrieval in a multimedia database is necessarily an interactive process because the user cannot formulate his multimedia query. Finally, since no available technique to handle the objects as active objects is sufficient to provide access to the multimedia data, we have to combine the retrieval results for different representations.

These requirements can only be met if we extend the design of a conventional database system. We introduced an architecture that can provide access to multimedia data. Further research is necessary to investigate how the probabilistic text retrieval model can be applied to the retrieval of multimedia objects. Our current research investigates how to enhance this framework for the purpose of multimedia retrieval. Although many aspects of multimedia databases have been studied, we still have a long road to take before multimedia database technology can realize its promises for human activity.

References

[Barrow, 1995] Barrow, J. (1995). *The artful universe.* Little, Brown and company.

[Bertino et al., 1995] Bertino, E., Catania, B., and Ferrari, E. (1995). Research issues in multimedia query processing. In *Advanced Course: Multimedia Databases in Perspective*, CTIT Report 95-25, pages 279–314. Center for Telematics and Information Technology of the University of Twente.

[Callan et al., 1992] Callan, J., Croft, W., and Harding, S. (1992). The INQUERY retrieval system. In *Proceedings of the 3rd international conference on database and expert systems applications*, pages 78–83.

[de Vries, 1996] de Vries, A. (1996). Television information filtering through speech recognition. In *Interactive Distributed Multimedia Systems and Services*, pages 59–69, Berlin, Germany. Springer.

[Eaglestone and Vertegaal, 1994] Eaglestone, B. and Vertegaal, R. (1994). Intuitive human interfaces for an audio-database. In *Proceedings of the Second International Workshop on Interfaces to Database Systems (IDS94).* Lancaster University.

[Edwards, 1993] Edwards, B. (1993). *Drawing on the right side of the brain.* Harper Collins Publishers, London.

[Faloutsos, 1996] Faloutsos, C. (1996). *Searching multimedia databases by content.* Kluwer Academic Publishers, Boston/Dordrecht/London.

[firefly, 1996] firefly (1996). http://www.ffly.com/.

[Iaccino, 1993] Iaccino, J. F. (1993). *Left brain-right brain differences.* Lawrence Erlbaum Associates, Publishers.

[Miller and Johnson-Laird, 1976] Miller, G. and Johnson-Laird, P. (1976). *Language and perception.* Cambridge university press.

[Niblack et al., 1993] Niblack, W., Barber, R., Equitz, W., Flickner, M., Glasman, E., Petkovic, D., Yanker, P., and Faloutsos, C. (1993). The QBIC project: querying images by content using color, texture and shape. Technical Report RJ 9203, IBM Research Division.

[Pearl, 1989] Pearl, J. (1989). *Probabilistic reasoning in intelligent systems.* Morgan Kaufmann, California.

[Salton, 1989] Salton, G. (1989). *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer.* Addison Wesley Publishing.

[Shardanand and Maes, 1995] Shardanand, U. and Maes, P. (1995). Social information filtering: Algorithms for automating "word of mouth". In *CHI'95 Proceedings*, Denver, CO, USA.

[van Rijsbergen, 1979] van Rijsbergen, C. (1979). *Information retrieval.* Butterworths, London, 2nd edition.

[Wactlar et al., 1996] Wactlar, H., Kanade, T., Smith, M., and Stevens, S. (1996). Intelligent access to digital video: The informedia project. *IEEE Computer*, 29(5).

[Wold et al., 1996] Wold, E., Blum, T., Keisler, D., and Wheaton, J. (1996). Content-based classification, search, and retrieval of audio. *IEEE Multimedia*, 3(3).