# The Psychology of Multimedia Databases

*Mark G.L.M. van Doorn*
University of Twente/Philips Research
mark.van.doorn@philips.com

*Arjen P. de Vries*
University of Twente/CWI
arjen@acm.org

**ABSTRACT**
Multimedia information retrieval in digital libraries is a difficult task for computers in general. Humans on the other hand are experts in perception, concept representation, knowledge organization and memory retrieval. Cognitive psychology and science describe how cognition works in humans, but can offer valuable clues to information retrieval researchers as well. Cognitive psychologists view the human mind as a general-purpose symbol-processing system that interacts with the world. A multimedia information retrieval system can also be regarded as a symbol-processing system that interacts with the environment. Its underlying information retrieval model can be seen as a cognitive framework that describes how the various aspects of cognition are related to each other. In this paper we describe the design and implementation of a combined text/image retrieval system (as an example of a multimedia retrieval system) that is inspired by cognitive theories such as Paivio's dual coding theory and Marr's theory of perception. User interaction and an automatically created thesaurus that maps text concepts and internal image concept representations, generated by various feature extraction algorithms, improve the query formulation process of the image retrieval system. Unlike most "multimedia databases" found in literature, this image retrieval system uses the the functionality provided by an extensible multimedia DBMS that itself is part of an open distributed environment.

**KEYWORDS:** Cognitive psychology and information retrieval, user and domain knowledge in query formulation, Paivio's dual coding theory, Marr's theory of perception

**Introduction**
Disclosure of multimedia content is becoming increasingly important as digital libraries grow quickly in both size and availability. Traditionally, access to multimedia has been through human-generated textual annotation. But annotation is costly in terms of both time and money and often subjective; how a person annotates differs from person to person and from time to time. Content-based multimedia retrieval which works on the perceptual signal itself avoids most of these problems but is unfortunately a difficult task for non-trivial application domains. Many content-based retrieval systems exists, but there is one system that surpasses them all – the human brain.

There are two reasons why cognitive psychology and science, that are concerned with topics like perception, learning, memory, language, emotion, concept formation and thinking from a human perspective, are also important in multimedia databases and content-based multimedia retrieval in particular. The first is that cognitive models of the human mind, which describe how the various aspects of cognition are related to each other, resemble information retrieval models: The human mind is a general-purpose symbol-processing system that interacts with the world, analogous to an information retrieval system that interacts with its world, different users with different information needs. Since this information processing system works so remarkably well, much time and effort can be saved in the search for some undiscovered, alien mechanism that might do the same. And secondly, by emulating human perception in multimedia databases one might automatically get a query processor that is better suited to capture and understand the goals, expectations, emotional state and other cognitive processes in the mind of the user during query formulation [23].

The remainder of this paper is organized as follows: First we discuss some important cognitive theories and show how they relate to methods and techniques, frequently applied in content-based multimedia retrieval and image retrieval in particular. Then we describe an extensible multimedia DBMS, the Mirror DBMS, that offers the functionality to build multimedia information retrieval systems. Next, we show how this functionality can be used to design and implement an image retrieval system [9]. The main contribution of this paper is making explicit the relationship to cognitive theories that

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
*Digital Libraries*, San Antonio, TX.
Copyright 2000 ACM 1-58113-231-X/00/0006...$5.00

inspired our work. We conclude this paper with some experimental results, conclusions and directions for future work.

### Cognitive psychology and content-based retrieval

Perception appears to be a simple and effortless process that we often take for granted. Closer examination shows that it is actually very complex: Many processes are involved in transforming and interpreting sensory input. Over 50% of our brain is estimated to be occupied by vision alone [13]. Marr's cognitive theory of perception (1983) [14] proposes that the processes involved in low-level vision produce a series of representations providing increasingly detailed information about the visual environment. Three kinds of representations are identified:

- The *primal sketch* provides a two-dimensional description of main light-intensity changes in the visual input including information about edges, contours and blobs.
- The *2 1/2-D sketch* incorporates a description of depth and orientation of visual surfaces making use of the information provided by color, texture, motion, and so on.
- The *3-D model* representation describes the shapes of objects in three dimensions that their relative positions in a way that is independent of the observer's viewpoint.

Object detection, the transformation of the 2 1/2-D representation to the 3-D representation, must cope with overlapping objects in the visual environment, a wide range of viewing distances and orientations and the fact that an object is a representation of concept, say this paper in front of you. Object recognition in Marr's 3-D model involves matching the 3-D representation constructed from the visual stimulus against a catalogue of previously learned 3-D models stored in memory. There is a significant amount of evidence from cognitive neuropsychology for Marr's theory of perception [10]. Many image retrieval systems seem influenced by Marr's theory of perception: Systems such as QBIC [1] and Photobook/FourEyes [16] use or combine the information provided by different feature extraction models to improve retrieval since there appears to be no single best feature extraction model. Moreover, the performance of feature extraction techniques depends on the data and type of query [17]. Image retrieval systems like VisualSEEK [15] and BlobWorld [4] use spatial information in combination with surface texture and color information to represent images. To a certain extent, this information is similar to the information contained in the primal and 2 1/2-D representations. However, depth perception, important in object recognition, is not implemented in most image retrieval systems.

Concept representation is important in higher cognitive processes like reasoning and knowledge organization. Paivio's dual coding theory (1971) (see figure 1) tries to model *mental imagery*, i.e. how we represent images and words in our mind. Basically, Paivio suggests that two independent but interconnected symbolic systems underly human cognition: a non-verbal and a verbal system each consisting of basic representation units called *imagens* and *logogens* respectively. The logogens system can be compared with a textual thesaurus used in text retrieval to improve query formulation (recall) by adding related words to the words in the original query. The imagens system on the other hand can be regarded as a visual thesaurus [22] that contains the non-verbal, internal image concept representations and their relations. The interconnections between these two symbolic systems would be analogous to a thesaurus that maps verbal, textual concept representations to non-verbal, image concept representations and vice-versa.

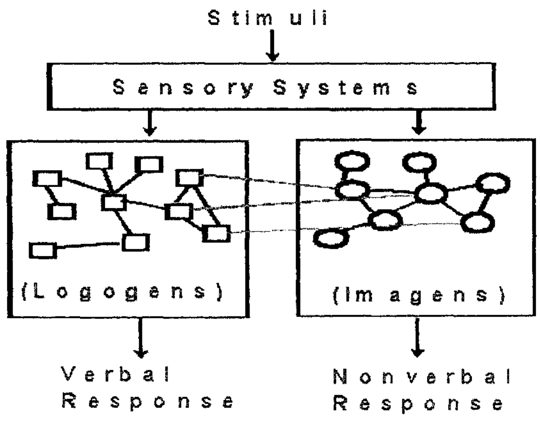


**Figure 1: Paivio's dual coding theory (1971)**

On the issue of the broad distinctions between verbal and non-verbal processing in the brain, dual-coding theory is moderately successful. Paivio further found that pictures were remembered, in both free-recall and recognition tasks, more readily than words, suggesting that the image code is mnemonically superior to the verbal code.

Paivio's theory seems more detailed than general information processing models like Wicken's model (1992) with regard to modeling long-term memory and the use of thesauri in concept reasoning. Unfortunately, dual-coding theory says little about the structure of imagery [10], the internal organization of knowledge. We need knowledge about things in order to behave and act in the world. It is important that knowledge is represented in natural concept classes and stored in an economic and informative way at the same time. Otherwise we would not be able to generalize or learn well over similar but previously unseen concepts.

Template cognitive theorists argue that stimuli are matched against miniature copies or templates of previously presented patterns. This is also the approach taken by most Query By Example (QBE) retrieval systems [1]. But given there is an almost infinite amount of templates to handle all previously encountered stimuli, template theories seem inadequate to account for the versatility of perceptual processing. Prototype concept theories view concepts as being organized around prototypes, expressed as clusters of attribute values similar to clusters of feature vectors in content-based multimedia information retrieval. Prototype theories can account for gradients of typicality, fuzzy boundaries and different levels of abstraction. There is a significant amount of empirical evidence for prototype theories [10], indicating that perhaps feature comparison in content-based multimedia retrieval systems should also take place at the concept or feature cluster level instead of at the instance level. Unfortunately, both template and prototype theories are silent about the knowledge people have about relations between properties. Schank's schema theory (1972) acknowledges this and encodes the temporal and causal structure of the world in knowledge schemata which are called frames or scripts [24]. However, little is known about the way how we form schema-like structures. There is empirical evidence that some of the most crucial steps in mental growth in humans are based not simply on acquiring new skills, but on acquiring new administrative ways to use what one already knows; this is known as Papert's Principle [18]. These more complex structures improve our ability to compare previously unseen, unknown concepts with existing ones (learning by analogy) and perhaps also our problem solving capabilities.

### Mirror DBMS architecture

Multimedia databases need a database management system (DBMS) to handle large amounts of data efficiently and many users simultaneously. DBMSs traditionally offer a different kind of retrieval than information retrieval (IR) systems [27]: DBMSs are very good at handling large amounts of structured data but do not adequately support searching on content. IR systems on the other hand are build to cope with incomplete, natural language queries and partial or best-match searches on the content of the data but lack the support for structured data. Multimedia databases must support both data and information retrieval to support a wide variety of possible queries. Besides extra functionality, there is another, architectural, reason for an integrated DB/IR system for multimedia databases: (relational) DBMSs deal with issues like persistent, efficient storage of data, concurrency control and transaction management and are optimized for handling large amounts of data. By building IR systems on top of a DBMS, researchers and developers of IR systems can concentrate on the retrieval models and let the DBMS take care of the data management.

The Mirror DBMS is a research database system that offers exactly this kind of integration [7]: The Mirror DBMS uses an extensible structural object-oriented logical data model and query algebra, the MOA object algebra [2] that is mapped on a binary relational physical database (Monet). This separation of the logical object-oriented data model and the physical data model brings the notion of *physical data independence* to the world of object-oriented databases and provides an excellent basis for algebraic query optimization. The Mirror DBMS extends the MOA object algebra by defining new structures and operations on these structures for information retrieval that can be used in combination with the existing basic structures to create powerful queries that can manage both the logical and layout structure of multimedia documents [6]. This way the Mirror DBMS separates the multimedia retrieval functionality from the actual multimedia retrieval application.

Since digital libraries typically involve several players with conflicting wishes and needs (content providers, users, access providers) and many simultaneous users, a single database would soon become a bottleneck in a large digital library setting and difficult to maintain. An open, distributed environment avoids many of these problems. New components can easily be added or removed without taking the entire system down and could also balance workload much more efficiently [8]. The Mirror DBMS is therefore part of an open distributed architecture consisting of several components such as daemons which can perform all sorts of complex tasks, end-user devices, media-content servers (typically webservers) and one or more meta-data databases or search engines (the actual Mirror DBMS), all connected via a common software bus (CORBA).

### Information retrieval model

In order to build an image retrieval system with the functionality the Mirror DBMS provides, one must implement an information retrieval model. According to [28] an information retrieval model consists of three fundamental parts: An appropriate scheme to represent documents and queries, a ranking function which determines to which extent a document is relevant to a query and a query formulation module. Basically, an information retrieval model describes how cognition works in an information retrieval system: The three fundamental parts relate to various aspects of human cognition such as visual perception, concept representation, knowledge organization and memory. Models of the human mind must address how these various aspects relate and influence each other. As an information retrieval model describes the same for information retrieval systems, we might regard it as a model of the *mind* of a system that interacts with its world, different users with different needs. This argues for an information retrieval system that is highly adaptive and capable of understanding the cognitive processes in the minds of many different users, each in a different cognitive state.

### Representation scheme

To model the representation scheme, i.e. computer perception and concept representation, the MOA object algebra has been extended with the DOCREP structure which represents the document contents (i.e. its words and their statistics, frequencies). Together with standard MOA structures such as SET and TUPLE this provides enough possibilities to model rich, complex, hierarchical document collections. The DOCREP structure has been designed for text retrieval, but can also be applied to images if clusters (groups) of similar feature vectors are treated as words in text retrieval like in [25]: These clusters then become the basic blocks of "meaning" or concept representations in multimedia information retrieval, similar to words in text retrieval, which are already concept representations.

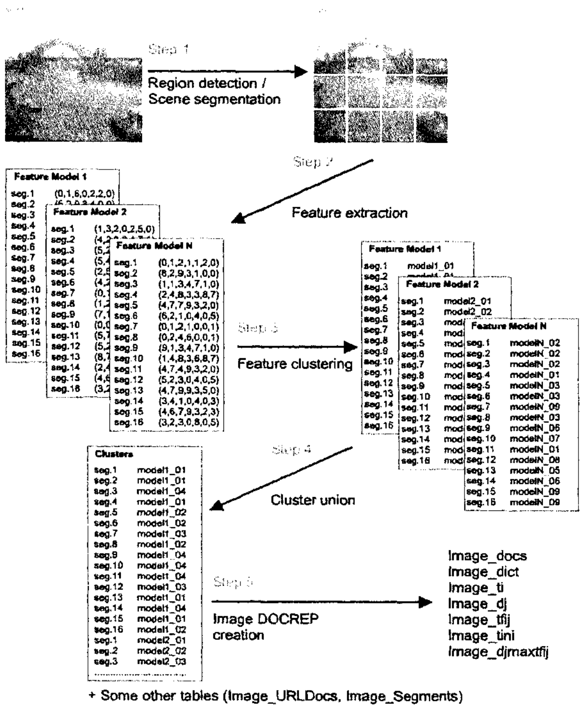


**Figure 2: Creating an image DOCREP structure**

Figure 2 shows how a DOCREP structure can be created from a collection of images. Many of these steps are implemented in daemons, software agents that provide additional functionality such as image collection from the web, segmentation and feature extraction. These daemons play an important role in the creation (indexing) of the meta-data database which provides content-based access to the multimedia documents stored on different multimedia content servers.

First, the images of the digital library are collected by a simple web robot. Another daemon then segments the images using a simple grid segmentation algorithm (step 1). The reason for a segmentation stage is the observation that most images like text documents typically represent multiple concepts. Grid segmentation is a rough approximation of region detection in Marr's primal sketch representation. The next stage (step 2) involves feature extraction on the individual image segments and is handled by several feature extraction daemons[1]. Like in FourEyes [17] the feature vectors of multiple feature models are calculated per image segment because there is no single best feature model. The feature extraction stage adds color and texture information to the individual image segments and belongs in the 2 1/2-D sketch of Marr's theory of perception. Each feature space is then clustered (step 3) using an unsupervised clustering algorithm[2]. These clustered feature spaces are similar to the imagens system proposed by Paivio's in his dual-coding theory (see figure 1). These clusters are then concatenated (step 4) and associated to the image they belong to. The final step creates the DOCREP structure that gives a content-based representation of the image. The following MOA structure shows how a collection of images is represented as a set of images with an URL and a textual[3] and internal DOCREP document representation:

```
define ImgLib as
SET<
  TUPLE<
    Atomic<URL>:source,
    DOCREP<Text>:annotation,
    DOCREP<Image>:content
  >
>;
```

### Ranking function

To determine which documents are relevant to a particular query, some sort of ranking function must be implemented. This is analogous to the process of memory retrieval in cognitive psychology as the input from the sensory systems (the user) is processed and compared with past experience (images in the document collection). MOA has been extended with a structure

[1] At this moment four texture and two color histogram feature extraction daemons have been built, using the reference implementations of the MeasTex package.

[2] We use AutoClass [5].

[3] The creation of a textual DOCREP structure is much easier because the words in the textual annotation are already concept representations and is therefore not explained here.

to model the collection statistics, DCSTAT and an operation on a DOCREP structure, getBL, to calculate the degree or belief to which each document in the collection is relevant to the given query terms (image feature clusters). The underlying Mirror retrieval model is an adaptation of the successful text information system INQUERY [3] that is based on probabilistic theory of evidential reasoning using Bayesian inference networks. The ranking function (the results are not sorted here) of a content-based query in MOA then becomes[4]:

```
map[TUPLE<name,sum(getBL(content, query, stats))>]
( ImgLib );
```

### Query formulation and knowledge

The feature clusters may be equivalent to words in text retrieval but these clusters are *internal* image concept representations (such as 'gabor_21') that are not suitable for interaction with the user: A mechanism is needed to provide a mapping between the concepts in the user's mind and these internal, machine-level clusters. A thesaurus that models the relationships between text concepts and internal image clusters can provide this mapping. It also enables us to find the best feature cluster (and feature model) for a particular query, similar to FourEyes' grouping weighting stage [16] which places prior weights on earlier generated groupings, clusters. This concept mapping thesaurus can be regarded as a partial implementation of Paivio's dual coding theory, more specifically the interconnections between the logogens and imagens subsystems in figure 1.

To create an association thesaurus that maps concepts to image clusters, a partially[5] annotated image collection is needed: The words in the textual annotations can then be associated with the clusters in the image representations. If a word and a particular cluster co-occur frequently in the document collection, the word is attached as a label to this cluster and added to the thesaurus. In other words, the available textual annotation is used to label or classify the unlabeled image content, like in [20] but applied to unlabeled multimedia content. A similar type of statistical caption-picture co-occurrence has been applied in the MARIE project [21]. The association thesaurus can also be represented as a DOCREP structure except this time the documents are replaced by image clusters and the getBL operation is used to measure the degree to which a cluster is relevant to a particular concept[6]. This approach is taken from PhraseFinder [12] which also uses the probabilistic model of the document collection for reasoning in the concept space.

Unfortunately, this mapping is not perfect: the co-occurrence thesaurus generates a lot of false relations, due to imperfect feature extraction models, clustering techniques and the simplicity of the co-occurrence method itself. To overcome this problem, the system should be able to adapt and improve itself. Like the human mind it should try to get a better understanding of its world, the individual users: It should learn from user interaction or relevance feedback across sessions. As explained in [19] for example, each learner needs an inductive bias to generalize beyond observed training examples. The concept thesaurus used during query formulation functions as an inductive bias for the document collection: It provides a basis for choosing good generalizations or in this case natural mappings of a concept to a set of internal representations (image feature clusters) to retrieve documents from the document collection that are relevant to the user's information request. When the number of examples is large the need for an inductive bias is low; inductive learning methods like neural networks and decision trees then have enough training data to achieve a high generalization accuracy. But the low-bias approach is not suitable for user-interaction since relevance feedback is costly in terms of the user's time. To overcome this problem the learning component starts with an initially created but imperfect concept domain model (an automatically created set of associations between concepts and feature clusters) that can be improved by user interaction[7]. The learner can not only change the vocabulary of the document collection, i.e. which feature clusters belong to a document (by modifying the inference network of the document collection); it can just as easily modify the concept language (weightings on these feature clusters) by changing the thesaurus inference network. This way, clusters in the thesaurus which are satisfactory this time for a particular query can be selected earlier the next time; the learner learns faster from previous interaction because it can make analogies from past experience [17]. This combination creates a continuous learner capable of modifying its concept language and transferring training from old problems to new ones [16].

Query formulation now takes place as follows (see figure 3): First, the user enters a textual query (top left). This query is used to consult the thesaurus which returns the most relevant clusters given this initial query. Next, the result of this query is used to generate a new query based on the image content (middle left) that is send to the meta-data database (bottom left). The results of this query are shown to the user (upper right). Relevance feedback of the user can be used to improve the query by

[4]The specifications of query and stats have been omitted.

[5]Annotating just a few images instead of the whole collection can save a lot of time and money.

[6]The thesaurus is implemented as a separate daemon but uses the same code as the meta-data database.

[7]Long-term adaptivity is not implemented at this time.

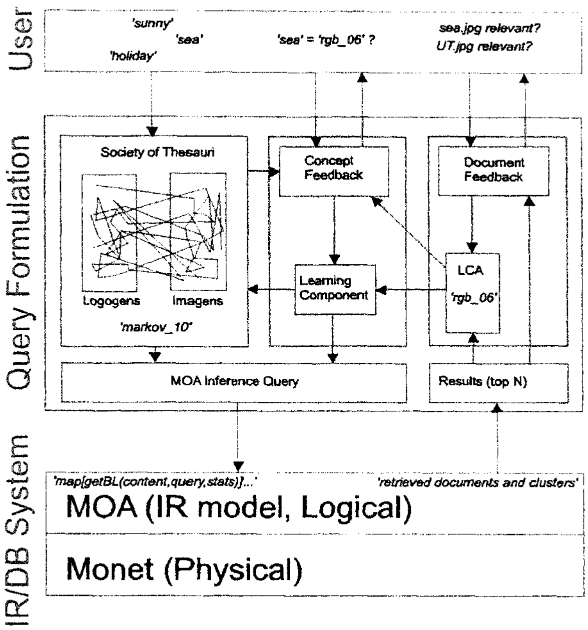


**Figure 3: Query formulation**

creating a new content-based query of the top N most occurring clusters in the relevant document set (LCA component) like in [11] is done for text. The middle column (concept feedback and learning component) in figure 3 has not been implemented at this moment. For more implementation details see [26].

### Evaluation

The Mirror DBMS is still very experimental and therefore no real evaluation with users has taken place. As a result of this no strong conclusions about the current implementation of the retrieval model inspired by cognitive theories can be made. However, some small-scale experiments have been performed, to illustrate the weaknesses of the current retrieval model and stress the importance of interactive learning from the user and the results of these experiments will be described here.

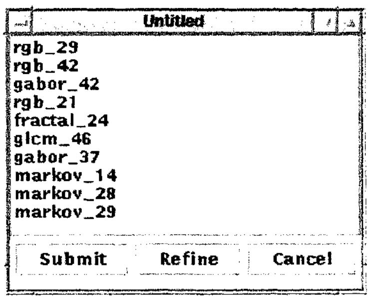


**Figure 4: Image clusters related to the concept "street"**

The image collection used is the 'BT' collection, also used in [16], and consists of 99 images. A small number, but the collection is diverse (pictures ranging from faces to landscapes) and there is little available training data for thesaurus construction so it is still quite challenging. The images have been annotated manually, and the annotations are used to construct a co-occurrence thesaurus. For example, figure 4 shows the image feature clusters co-occurring frequently with the text concept "street" (obviously, this representation is not intended to be shown to end-users).

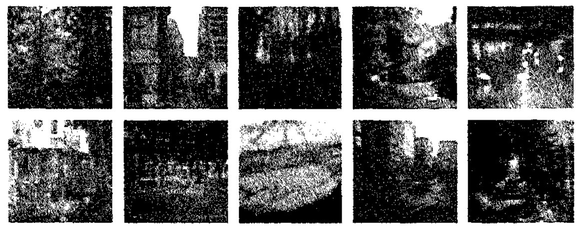

**Figure 5: Top 10 for the query "tree, trees, forest"**

Figure 5 shows the results of a textual query consisting of the words "tree", "trees" and "forest", expanded using the automatically constructed concept mapping thesaurus. Figure 6 and 7 show images that contain clusters that the thesaurus associated with these words. Closer examination of the query formulation process shows that two classes of errors can be recognized: clusters with little perceptual meaning and clusters with a wrong label. In the first class of errors, AutoClass finds clusters of little semantic value. Figure 6 shows an example of this class of errors. Labeling errors occur when the same text concepts and image clusters co-occur frequently in he text collection; image representations can then be erroneously labeled during the construction of the association thesaurus because image clusters are associated with the wrong text concepts as in figure 7 for example.

**Figure 6: Images that contain clusters 'fractal_23'**

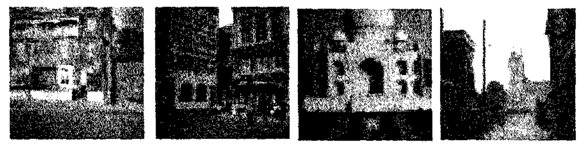

**Figure 7: Images that contain clusters 'gabor_20'**

Fortunately, in other cases, the automatic mapping from text concepts to image clusters is working. As an example consider figure 8 in which the image cluster 'glcm_47' is correctly associated with the high-level concept represented by the text keywords "tree", "trees" and "forest".

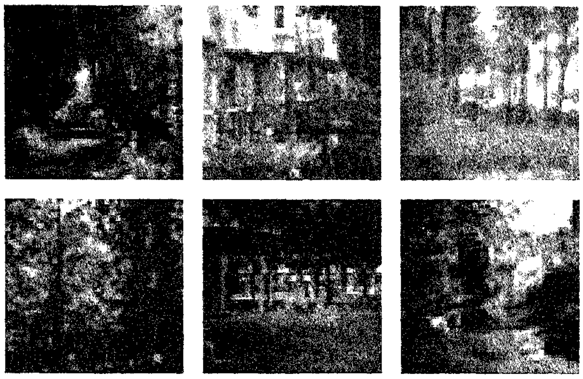

**Figure 8: Images that contain clusters 'glcm_47'**

**Figure 9: The top 6 images retrieved after relevance feedback on the first and fourth image in figure 5**

User interaction (i.e. relevance judgments of the user) allows the system to improve itself by removing image clusters such as 'gabor_20' and 'fractal_23' and adding image clusters like 'glcm_47'. The system could use this knowledge to update its domain knowledge stored in the thesaurus and the document collection itself and thus learn from interaction within *and* across sessions. However, such a continuous learner has not been implemented yet. The first impression of relevance feedback is that it really improves the content-based query processing in the prototype system. For example, if the first and fourth image of Figure 5 are given as positive examples, the system returns the images shown in figure 9. Also, it has proven possible to retrieve images with faces by giving positive relevance feedback on portraits, even though the current implementation does not have special features for this purpose.

## Conclusions

We believe that cognitive psychology and science, which are concerned with cognition from the human perspective, can also provide insight in the difficulties surrounding multimedia indexing and retrieval. In this paper we argued that a better understanding of human cognition may lead to better-performing content-based multimedia information retrieval systems for two reasons: First, the human mind viewed by cognitive psychologists as a general-purpose symbol-processing system that interacts with the world, is analogous to a multimedia information retrieval system that interacts with different users. The human mind is extremely good at this and is undoubtably the best-performing example of an "information retrieval" system in this sense. Secondly, by emulating human perception in multimedia databases one might automatically get an information retrieval system that is better suited to capture and understand the goals, expectations, emotional state and other cognitive processes in the mind of the user during query formulation. Expressing the user's information need in a machine-understandable way is exactly what query formulation in multimedia databases is all about!

An information retrieval model can be regarded as the cognitive framework for an information retrieval system: It describes how different cognitive theories are related to each other. In other words, an information retrieval model describes how the *mind* of the information retrieval system works and interacts with different users. An extensible multimedia DBMS, the Mirror DBMS, provides us with the information retrieval structures needed for concept and image representation and an operation for probabilistic inference on these structures, which allows us to retrieve earlier stored concept and image representations from memory. With this functionality we have implemented an image retrieval system inspired by cognitive theories.

Before images can be represented in a machine-understandable way, we need to extract information from the image signal. Marr's theory of perception suggests that vision involves a series of representations providing increasingly detailed information about the visual environment. We approximate Marr's theory in the following way: First we use a simple grid segmentation algorithm to segment the images, then we apply multiple feature extraction techniques on each segment to obtain multiple representations of each segment. The resulting feature vectors are grouped using an unsupervised clustering algorithm. Each image in the document collection is represented as a set of these clusters and stored in the Mirror DBMS in a so-called DOCREP structure. We use an automatically constructed co-occurrence thesaurus (on a partially annotated image collection) to map (text) concepts to image clusters. This thesaurus can be regarded as an partial implementation of Paivio's dual coding theory, a cognitive theory describing how we represent imagery in our mind. Furthermore, it allows the system to select feature clusters (and models) without explicit user-interaction; it assists in the query formulation process. Since the thesaurus is defined in the same information retrieval structures as the document collection, we can search the thesaurus or concept representation space using the same probabilistic inference technique that we use to search in the document collection: The retrieved concept representations or image clusters provide the query terms for a content-based search in the image representation or document collection space. The result of this search is a list of relevant images, ready to be shown to the user.

The concept and image representations stored in the thesaurus and image collection is imprecise and incomplete at first and changes over time as the goals of the user change, therefore user-interaction is necessary: Relevance feedback from the user is used to improve the initial query formulated by the thesaurus and could be used by a continuous learner to update its inductive bias, encapsulated in the thesaurus, and the document collection itself to improve the system across sessions. In Bayesian inference models this implies document collection and concept thesauri modification, but this has not been implemented yet.

The image retrieval system has not been evaluated against large, partially annotated image collections; but, most components in figure 3 have been implemented and the extra domain knowledge in the thesaurus is certainly better than random selection of feature clusters. Relevance feedback on the retrieved image set shows significant improvements on most queries, so we can expect good results if the thesaurus and image collection inference networks can be modified.

### Future work

Many improvements can be made to both the design and implementation of the system. First of all, the cognitive framework must be made more robust and scalable to larger, multi-user environments: Empirical results show that the simple association thesaurus does not scale to larger annotated image collections due to computational complexity of the current co-occurrence calculation algorithm. Furthermore, an implementation of the continuous learner, which has shown significant improvements in FourEyes, is needed to improve the system's performance not only within but also across sessions.

Other interesting future work concerns the combination of evidence from multiple modalities. Having multiple modalities could reduce the overall uncertainty in the formulated query and increase recall and precision in probabilistic multimedia information retrieval systems: If humans do not recognize an object by its visual appearance we can often identify it by using (one or more of) our other sensory input organs although most of the time we are probably not even aware of this.

We also need to do usability studies: Evaluation of the system against larger annotated image collections, testing the effect of a changing system on users and investigating the effects of different granularities of annotation on retrieval performance for different users are required before stronger assumptions about the current system and its underlying cognitive framework can be made.

Finally, we are looking at ways to integrate the cognitive model underlying our multimedia information retrieval system with intelligent user interfaces since they complement and are also likely to influence each other.


### Acknowledgements

We like to thank Annita N. Wilschut and J. Flokstra for the MOA kernel and Erik van het Hof for the open distributed environment. Finally, we would also like to thank the Monet team, Peter Boncz in particular, for all their support.


### REFERENCES


1. J. Ashley, J.L. Hafner, D.L. Lee, W. Niblack, and D. Petkovic. The Query By Image Content (QBIC) system. In M.J. Carey and D.A. Schneider, editors, *Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data*, page 475, San Jose, CA, May 22-25 1995.

2. P. Boncz, A.N. Wilschut, and M. Kersten. Flattening an object algebra to provide performance. In *Fourteenth International Conference on Data Engineering*, pages 568–577, Orlando, Florida, Februari 1998.

3. J.P. Callan, W.B. Croft, and S.M. Harding. The INQUERY retrieval system. University of Massachusetts.

4. C. Carson, M. Thomas, and S. Belongie. Blobworld: A system for region-based image indexing and retrieval. *Third International Conference on Visual Information Systems*, June 1999.

5. P. Cheeseman and J. Schutz. *Bayesian Classification (AutoClass): Theory and Results*. NASA. http://ic-www.arc.nasa.gov/ic/projects/bayes-group/autoclass/.

6. A.P. de Vries. Mirror: Multimedia query processing in extensible databases. In D. Hiemstra, F. de Jong, and K. Netter, editors, *Proceedings Twente Workshop on Language Technology (TWLT14): Language Technology in Multimedia Information Retrieval*, volume 14, pages 37–47, Enschede, the Netherlands, December 1998.

7. A.P. de Vries. *Content and multimedia database management systems*. PhD thesis, University of

Twente, Enschede, The Netherlands, December 1999.

8. A.P. de Vries, B. Eberman, and D. Kovalcin. The design and implementation of an infrastructure for multimedia digital libraries. In *Proceedings of the 1998 International Database Engineering & Applications Symposium*, pages 103–110, Cardiff, UK, July 1998.

9. A.P. de Vries, M.G.L.M. van Doorn, H.M. Blanken, and P.M.G. Apers. The Mirror MMDBMS architecture. In *Proceedings of 25th International Conference on Very Large Databases (VLDB '99*, pages 758–761, Edinburgh, Scotland, UK, September 1999. Technical demo.

10. M.W. Eysenck and M.T. Keane. *Cognitive psychology: a student's handbook.* Lawrence Erlbaum Associates, 3rd edition, 1995.

11. D. Haines and W.B. Croft. Relevance feedback and inference networks. In *Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2–11.

12. Y. Jing and W.B. Croft. An association thesaurus for information retrieval. Technical report, University of Massachusetts, 1994.

13. S.M. Kosslyn and D.N. Osherson, editors. *Visual cognition: An invitation to cognitive science*, volume 2. MIT Press, Cambridge, MA, 1995.

14. D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* W H Freeman & Co, 1983.

15. M.T. Maybury, editor. *Intelligent Multimedia Information Retrieval.* MIT Press, Cambridge, MA, 1997.

16. T.P. Minka. An image database browser that learns from user interaction. Master's thesis, MIT Media Laboratory Perceptual Computing Section, 1996.

17. T.P. Minka and R.W. Picard. Interactive learning using a "society" of models. Technical Report 349, MIT Media Laboratory Perceptual Computing Section, 1997. Submitted to Special Issue of Pattern Recognition on Image Databases: Classification and Retrieval.

18. M. Minksy. *Society of Mind.* Simon & Schuster, 1985.

19. T.M. Mitchell. *Machine Learning*, chapter Combining inductive and analytical learning. McGraw-Hill Companies Inc., 1997.

20. K. Nigam, A. McCallum, S. Thrun, and T. Mitchell. Learning to classify text from labeled and unlabeled documents. *Machine Learning Journal*, 1999.

21. N.C. Rowe. Retrieving captioned pictures using statistical correlations and a theory of caption-picture co-reference. In *Fourth Annual Symposium on Document Analysis and Retrieval.* Las Vegas, April 1995.

22. R.W.Picard. Toward a visual thesaurus. Technical Report 358, MIT Media Laboratory Perceptual Computing Section, 1995.

23. R.W.Picard. *Affective Computing.* MIT Press, 1997.

24. R. Schank. *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures.* Lawrence Erlbaum Associates, 1977.

25. D. McG. Squire, W. Muller, H. Muller, and J. Raki. Content-based query of image databases, inspirations from text retrieval: inverted files, frequency-based weights and relevance feedback. In *The 11th Scandinavian Conf. on Image Analysis*, pages 7–11, Kangerlussuaq, Greenland, June 1999.

26. M.G.L.M. van Doorn. Thesauri and the Mirror retrieval model. Master's thesis, University of Twente, July 1999.

27. C.J. van Rijsbergen. *Information Retrieval.* Butterworths, London, 2nd edition, 1979. See also: http://www.dcs.glasgow.ac.uk/Keith/Preface.html.

28. S.K.M. Wong and Y.Y. Yao. On modeling information retrieval with probabilistic inference. In *ACM Transactions on Information Systems*, volume 13, pages 38–68, 1995.