Extending Image Retrieval Systems with a Thesaurus for Shapes

Master Thesis
Lars-Jacob Hove
Institute for Information and Media Sciences
University of Bergen
Lars.Hove@student.uib.no
October 12th, 2004

Abstract

Successful retrieval of relevant images from large-scale image collections is one of the current problems in the field of data management. In this paper, I propose a system which combines techniques from Computer Vision, such as feature extraction, with a thesaurus for objects / shapes. A brief presentation of the problem area is given, along with a presentation and test results from a prototype system, VORTEX. The VORTEX system has been compared to a retrieval system based on low-level image features alone, and is able to achieve a significantly higher degree of image recall.

Keywords

Information retrieval, Image retrieval, CBIR, thesaurus, feature extraction, Image databases

Introduction

During the last decade we have seen a rapid increase in the size of digital image collections. As the computational power of both hardware and software have increased, the ability to store more complex data types in databases, such as images, has been drastically improved. These new media types offer other challenges, and demand different treatment than text, in order to be useful in databases. Research in this area started in the 70's, based in traditional information retrieval, and is today both an active and important field in information and data management (Huang and Rui 1999).

There have been two distinct research communities driving research in image retrieval forward. The first has its roots in traditional information retrieval, using text based techniques for indexing and retrieval of images. The second major area is the field of computer vision, which uses methods from computer science to analyze and index images based on their visual content, such as color, texture or shape (ibid).

In the recent years, it has been suggested that a combination of techniques from both areas might yield better results than either approach alone (Santini and Jain 1998). In my master thesis at the Institute for Information Science and Media Studies at the University of Bergen, I propose a tool that aims to improve Image Retrieval; a *shape thesaurus*. It aims to draw upon the strengths of structural feature extraction, combined with a tried and tested technique from Information Retrieval; the Thesaurus.

Information Retrieval and Computer Vision

Information retrieval – semantic understanding

Indexing and retrieval of images using techniques and methods from traditional information retrieval has *semantic understanding* as one of its main strengths. It relies on textual description of images, mostly using keywords or free text. This has high expressive power; it can be used to describe most aspects of image content. In addition, the process of searching and retrieval can be automated by a wide range of existing text retrieval software. However, it has been shown that there are three major difficulties to this approach; subjectivity, volume, and explicability.

Textual description and annotation is mainly a manual process. The combination of rich image content and differences in human perception makes it possible for two individuals to have diverging interpretations of the same image. As a result, the description is prone to be subjective and incomplete. In addition, the use of different words by the indexer and the searcher, such as synonyms or general / specific terms to describe similar images makes retrieval based on annotation even more difficult. (Santini and Jain 1998; Eakins and Graham 1999; Huang and Rui 1999). This points at an important distinction between indexing and retrieval of images using these techniques; while it is relatively easy to describe and index images, the search results depend on the enquirer's knowledge of the indexing terms used. (Lu 1999) suggests using domain knowledge or an extended thesaurus might help alleviate these problems.

The second problem, the potentially large amount of images in a collection, can make textual description and annotation a tedious and very time consuming process, with indexing times quoted up to as much as 40 minutes per image. (Eakins and Graham 1999; Huang and Rui 1999; Lu 1999).

Finally, while text based description has a high expressive power, there are some limitations when dealing with objects that are visible in nature. Some structural image characteristics are difficult to describe with words. For example, although we have a set of terms describing the different colours, none of these terms are exact. Every colour has a broad range of different shades and intensities. Although most people are able to differentiate between two different shades, it is difficult to express the differences verbally without using fuzzy terms like "more" or "less" red. We call this the problem of explicability.

Computer vision – Automated Processes

Another approach to retrieval from image databases is rooted in the field of computer vision. As a response to the difficulties large scaled image collections posed to traditional techniques, *Content based Image Retrieval,* or CBIR, was proposed. Rather than relying on (manual) annotation and textual descriptions, CBIR systems use automatic feature extraction for indexing and retrieval of images based on image content. The features used for retrieval can either be primitive or semantic, but the extraction process must be predominantly automatic. Current CBIR systems operate on *primitive image features*, such as color, texture, shape and spatial structures (Eakins and Graham 1999).

The basics of CBIR systems are thoroughly described in literature, (Lu 1999; Li and Kuo 2002) and these are not elaborated further here. However, I will highlight what is considered one of the greatest challenges to CBIR systems; *bridging the semantic gap*.

While most current commercial solutions only operate on the lowest level, using the extracted primitive features as basis for a similarity search, we need to perform retrieval based on higher level content, such as objects, scenes and activities. Recent research has shown that while primitive features work well for application areas where information is embedded in the structure of the image, such as fingerprint identification and medical diagnostics. However, they do not provide adequate support for more general application areas. Much of the concurrent research is aimed at bridging the gap between low level features and semantic content, dubbed the "semantic gap" (Eakins and Graham 1999; Colombo and Del Bimbo 2002).

(Santini and Jain 1998) illustrate the above mentioned problem. A CBIR system that solely depends on extraction and comparing of primitive features has no understanding of the image's semantic contents. For example; the user submits an image of a dark haired woman standing in front of a white wall. The retrieval system finds several images of portals and doors in a white wall. Even though the visual features of the two images might be comparable, there is no semantic likeness. The problem arises because images can be similar at different levels; in this case the color, shape and texture. The CBIR system is able to identify these lower-level similarities, but fails to recognize similarities in semantic content.

Combining the Best of Two Worlds – bridging the semantic gap

If the two approaches could be combined, it seems likely that the strength of one field could offset weaknesses of the other, thus bridging the semantic gap. (Huang and Rui 1999) suggest that an interdisciplinary research effort is required in order to build a successful Image Database System. It is suggested that CBIR not is a replacement of text based retrieval, but a complementary component.

Huang and Rui (ibid:10) also suggest a possible integrated system architecture. The suggested architecture is based on three databases; a collection of images, a collection of visual features extracted from the images and a collection of

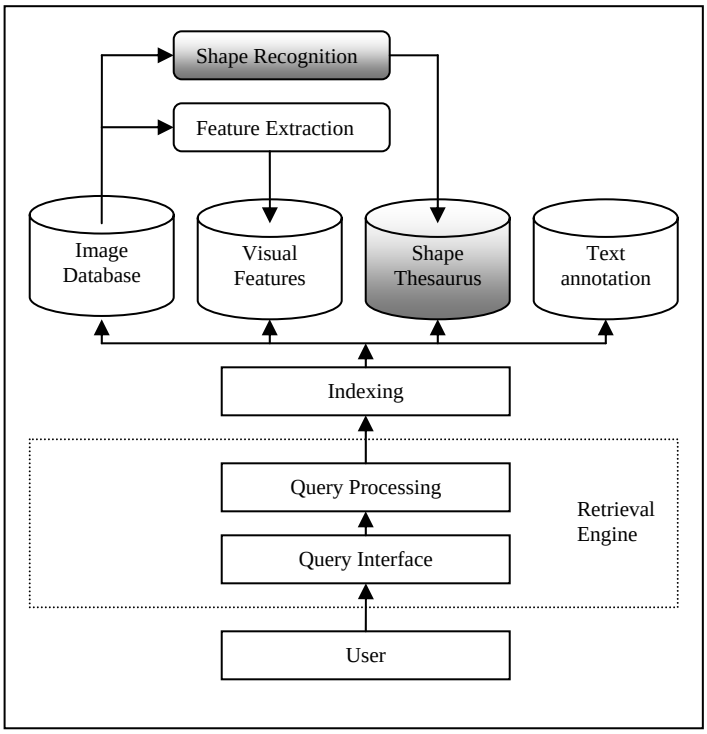

Figure 1: Architecture for a CBIR system extended with a thesaurus for shapes.

keywords and free text describing the images. This builds on the idea that text based annotation can work side by side with a CBIR system. It is suggested this is the only way satisfactory retrieval performance can be achieved. However, this does not address the problems of volume and subjectivity; the textual annotation is still a predominately manual process. I therefore propose an addition to this architecture, *a shape thesaurus*. Figure 1, above, shows an illustration of an image retrieval system with a thesaurus for shapes.

The Shape Thesaurus

As suggested in (Lu 1999), a thesaurus might alleviate the retrieval process by helping the retrieval system identify images annotated by keywords related to the terms specified in the query. The main purpose of a thesaurus is to give a standardized system of reference, for indexing and searching, to assist the user with locating the correct terms for query formulation and to provide classification hierarchies that allow the broadening and narrowing of the terms given by the user in the current query request (Foskett 1997). It is my belief that this principle might be translated into a thesaurus for shapes, thus linking shapes that are different in appearance but similar in semantic content together. It is hoped that this might result in more meaningful retrieval results than pure CBIR.

Recent research has suggested similar approaches to bridging the semantic gap. (Manjunath and Ma 2002) have reported a significant performance improvement when using self-organizing maps to cluster images based on extraction of texture feature vectors. Information links are created between images using code words and texture samples, thus creating a thesaurus of textures. (Dobie, Tansley et al. 1998) have presented an architecture that allows a multimedia thesaurus and agents to be combined with a content based hypermedia system, aiming to overcome the problems mentioned in this paper.

The following definition of a Shape Thesaurus is proposed:

> *(1) A precompiled list of t shapes representing important visual objects in a given domain of knowledge*
> *(2) statistical descriptors describing these shapes*
> *(3) a textual / semantic description of these shapes*
> *(4) for each shape, a set of related shapes.*

The definition presents the four major components of a shape thesaurus. The first (1) contains shape templates and examples describing important visual objects in a domain. The templates could be based on simple icons, sketches images representing the important visual objects in a domain, while the examples consist of clear images depicting these objects. This form for shape description is simplification of *deformable shape templates*, such as described in (Sclaroff and Liu 2001). Figure 2 shows excerpts from a list of shapes for a maritime thesaurus domain, with two templates

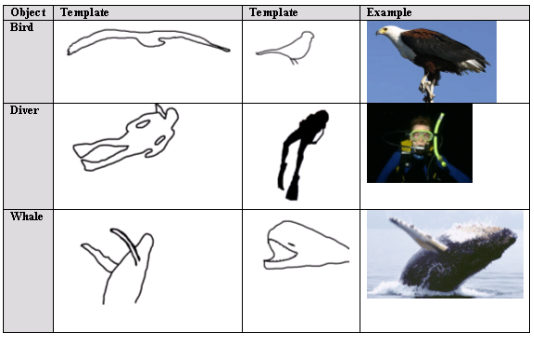

Figure 2: Excerpts from a shape thesaurus object list for a maritime domain.

and one example shown for each object. The templates were based on real images. In addition to important objects, some classification terms should be used as super types, such as “Mammal” and “Marine Mammal”.

Next, there must be a set of statistical shape descriptors (2). These are used for shape similarity comparison between the shape descriptors and both query images and an image collection.

Furthermore, there would be a textual / semantic description of these objects (3). Images differ from textual information in that images are essentially unstructured, since digitized images consist purely of arrays of pixel intensities, with no inherent meaning. Image data thus differ fundamentally from text, where the raw material, words stored as ascii character strings, has already been logically structured by the author. (Eakins and Graham 1999). It is therefore necessary to add a semantic description to the terms / items described by the thesaurus. It is then also possible for a user to use these textual descriptors as search criteria.

Finally, there is the set of relationships between the defined objects (4). A shape thesaurus consists of two relationship groups, *object relationships* and *shape relationships*. Figure 3 shows a presentation of relationships of the two groups.

The *object relationships*, is related to the visual objects the shapes represent. These relationships provide support for retrieval of images with content is semantically related to the query items requested, but not identical.

The *shape relationships*, is related to the shapes representing the visual objects.

Relationship	Description	Example
Object relationships		
Hierarchical	An object is a specialization or a generalization of its related object.	“dolphin fin” is a specialization of “fin”
Related	An object has an unspecified relationship to its related object	A “Whaling ship” is related to a “whale”
Shape Relationships		
Part-Of	A shape is a part of its related shape	A “Dolphin fin” is a part-of a “Dolphin”
Variant-of	A shape is a variant of its related shape.	The shape of a “jumping dolphin” is a variant-of a “swimming dolphin”.

Figure 3: Shape Thesaurus Relationships

It is believed that a thesaurus as described here will help alleviate some of the difficulties of content based image retrieval. Creating the structure for the thesaurus and defining relationships are trivial. The two major problems that must be overcome are developing representative shape descriptors, as well as mapping the thesaurus to the image collection. Various methods could be used for the shape descriptors, such as deformable shape templates, pattern-recognition algorithms, neural networks or other tried and tested techniques from computer vision.

Mapping the thesaurus to an image collection could either be done through manual assignment, or through object detection and recognition techniques. Manual assignment would only be applicable either for small image collections, or when adding a small amount of images to an existing collection. For large scale inserts, or when creating a new collection, identifying objects in the images will grow very cumbersome. Therefore, the

system has to provide some method of automatic identification of an image's contents. The techniques used in the thesaurus for representing and recognizing shapes could be applied here as well.

One possible approach to creating such a Shape Thesaurus, the VORTEX system, has been implemented. Other, possibly better approaches, have been evaluated and are further discussed below.

The VORTEX prototype

As described above, it is difficult to apply the discussed framework manually on a large scale image collection. There has to be an element of automation, and there are two major problems that have to be overcome. First, it is necessary to develop good and representative shape descriptors. These are fundamental to the system, as identification of thesaurus elements is reliant on these. Furthermore, a suitable CBIR mechanism must be provided for analyzing and comparing search images, the thesaurus representations and the image collection.

A prototype image retrieval system, Visual Object Retrieval – Thesaurus EXtension (VORTEX) has been implemented using Oracle 9i interMedia (O9i) as a foundation. O9i is an object-relational database management system and, with the interMedia package provides basic support for Content Based Image Retrieval. In addition O9i has support for both SQL3 and PL/SQL, making it both suitable for multimedia data, as well as relatively easy to expand the built in functionality (Nordbotten, 2004).

While the CBIR mechanisms provided by O9i are far from ideal, the limited time frame for this project and O9i's adaptability makes it preferable to develop a tool from scratch. In addition, it is believed that if the proposed framework can prove useful even with the limited tools provided in O9i, it is worthy of further studies.

The VORTEX system consists of 3 main components; an image collection, a thesaurus and a search engine, as illustrated in Figure 4. The image collection and the thesaurus are both made with object-relational object structures, while the search engine is made as a package of PL/SQL procedures.

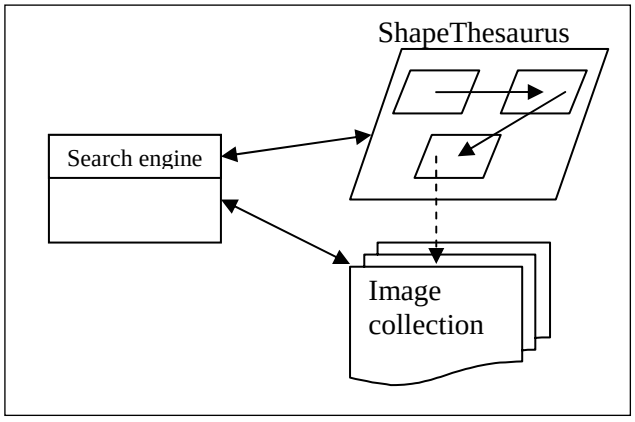

Figure 4: Architecture of the VORTEX system

The image collection is implemented as an object-relational structure, containing the image as a BLOB. In addition, it contains a signature of the image representing colour, shape, texture and spatial arrangement. This signature is used for comparison between images. In addition, the structure provides support for additional image indexing techniques, such as keywords.

The thesaurus is built much in the same way. It consists of a hierarchical list of shapes and objects, as well as relationships between them. In addition, each thesaurus element has a set of shapes, effectively creating a shape template. These are stored as BLOBS with signatures representing the statistical shape representations.

The thesaurus is based on a simple, hierarchical taxonomy of marine animals, as well as related subjects, and contains shapes and examples of the thesaurus objects. Figure 5 gives a description of the visual objects in the VORTEX shape-thesaurus. Actual shapes are represented through templates similar to the examples in figure 2, above. The shape relations indicated by the word in the parentheses, and has been implemented

Figure 5: Vortex Object List

The search engine consists of a package of procedures used for searching and comparing images. It basically takes a search image, generates a signature and compares this to the representations in the thesaurus, using characteristics derived from shape features alone. If and when a potential match is found, it searches the image collection for images similar to both the matching shape and the other representations of the thesaurus term. Next, the search image is compared to the image collection using all characteristics (Shape, colour, image and spatial placement). Finally, the best matches from both searches are ranked and presented.

Evaluation of VORTEX and the Shape Thesaurus

Test collection

The VORTEX prototype has been populated with images and a thesaurus based on a maritime scenario. A set of about 150 images and drawings depicting whales, dolphins, sharks, fish and other marine animals in different situations have been compiled for the project. The images range from very simple drawings of one animal, to complex images depicting animals in different situations. A small, customized collection has been chosen rather than an existing collection in order to have complete control over, as well as total knowledge of the images in the collection.

Testing the system

The VORTEX system will be compared to the standard CBIR functionality existing in O9i, and will be measured and compared using precision / recall. A small set of respondents were given queries expressed in text, and generated a set of 36 queries. After query specification, the respondents were given the complete set of images in the collection, and determined which of the images they found to be relevant for the queries they made.

The query images generated by the respondents were executed on both VORTEX and through the standard Oracle CBIR functionality. The results sets were compared to the list

of relevant images specified by the respondents, and used to measure recall and precision for the two systems. Single values representing recall and precision for each query was measured. Figure 6 shows an illustration of the recall and precision values for the two systems. 6A shows $Recall_{A/B}$, representing the difference in recall values between VORTEX and Oracle for each query. A negative value indicates a difference in favour of VORTEX. 6B shows similar values for precision.

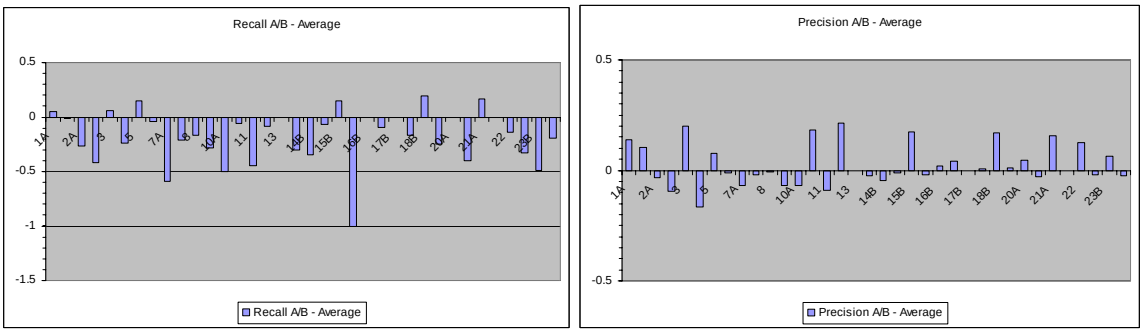

Figure 6. Recall and Precision for the 36 queries. A) shows recall, 6) shows precision.

Neither system achieved very high recall or precision rates. The results indicated that VORTEX was able to achieve higher recall. The difference was found to be significant at a 0.05 level of confidence with a Students-T test. The differences in precision were much less clear. Oracle managed to achieve slightly higher precision. However, the differences were not found to be significant.

Conclusions

Image retrieval with a shape thesaurus, as represented by the VORTEX prototype, has been found to achieve significantly better recall results over a system based on low level feature comparison alone. The VORTEX system is a limited implementation of a shape thesaurus, and it is difficult to draw any final conclusions based on this experiment. However, the findings are positive, which indicates that the proposed framework is worthy of further evaluation.

Future research

The ideas presented in this paper and in my master's thesis, are to be considered a pilot study for a larger and more thorough research project in a PhD thesis. Hopefully, the results will give an indication of the possibilities presented by using a shape / object based thesaurus in the context of an image retrieval system.

Results from a preliminary test of the systems indicate that the functionality provided in the framework described in this paper might provide useful. It is believed that the full experiment, which is in progress at the time of writing, will provide more conclusive evidence to this.

As mentioned, using Oracle 9i interMedia as a basis for the prototype has several limitations. For one, there is little control over the mechanisms used for identification of objects and comparisons between images. Different approaches to this, such as using a

neural network application, were examined during the development of the prototype. Due to the limited time frame provided by this project, further enquiries into these approaches had to be postponed. However, it is the belief of this author that better image analysis tools will provide better and more controlled results.

Currently, the VORTEX system does not have any functionality for refining itself; every search has to start at scratch. Possibly, a relevance-feedback between the system and the user will enable the system to learn “good” results, and thus create a link between the thesaurus elements and images.

Acknowledgements

This work has been done within the framework of an NFR project #148827/530, Virtual Exhibits on Demand. I would like to thank my advisor, associate professor Joan Nordbotten, for guidance and valuable discussions and insights.

References

Colombo, C. and A. Del Bimbo (2002). Visible Image Retrieval. Search and Retrieval of Digital Imagery. V. Castelli and L. D. Bergman, John Wiley & Sons, Inc**:** 11-33.

Dobie, M., R. Tansley, et al. (1998). A Flexible Architecture for Content and Concept Based Multimedia Information Exploration. Challenge of Image Retrieval, Newcastle.

Eakins, J. P. and M. E. Graham (1999). Content Based Image Retrieval: A report to the JISC Technology Applications Program. Newcastle, Inst. for Image Data Research, Univ. of Northumbria.

Foskett, D. J. (1997). Thesaurus. Readings in information retrieval. K. S. Jones and P. Willet, Morgan Kaufmann Publishers**:** 111-134.

Huang, T. S. and Y. Rui (1999). "Image Retrieval: Current Techniques, Promising Directions And Open Issues." Journal of Visual Communication and Image Representation **10**(4): 39-62.

Li, Y. and C.-C. J. Kuo (2002). Introduction to Content-Based Image Retrieval - Overview of key techniques. Image Databases: Search and retrieval of Digital Imagery. V. Castelli and L. D. Bergman, John Wiley & Sons, Inc**:** 261-284.

Lu, G. (1999). Image Indexing and Retrieval. Multimedia Database Management Systems, Artech House publishers**:** 131-161.

Lu, G. (1999). Multimedia Database Management Systems. Norwood, Artech House INC.

Manjunath, B. S. and W.-Y. Ma (2002). Texture Features for Image Retrieval. Image Databases: Search and Retrieval of Digital Imagery. V. Castelli and R. Baeza-Yates, John Wiley & Sons, Inc**:** 313-344.

Santini, S. and R. Jain (1998). Beyond Query by Example. ACM Multimedia'98. Bristol, UK, ACM**:** 345-350.

Sclaroff, S. and L. Liu (2001). "Deformable shape detection and description via model-based region grouping." IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(5): 475-789.