CHAPTER 2

LITERATURE SURVEY

2.1 INTRODUCTION TO IMAGE RETRIEVAL

Image retrieval has been an extremely active research area over the last 10 years, but first review articles on access methods in image databases appeared already in the early 80s (Chang and Kunii 1981). The following review articles of various years explain the state-of-the-art of the corresponding years and contain references to a large number of systems and descriptions of the technologies implemented. Enser (1995) gives an extensive description of image archives, various indexing methods and common searching tasks, using mostly text-based searches on annotated images. In (Gupta and Jain 1997), an overview of the research domain in 1997 is given and in (Rui et al 1997), the past, present and future of image retrieval are highlighted. In (Eakins and Graham 2000) an almost exhaustive overview of published systems is given and an evaluation of a subset of the systems is attempted in (Venters and Cooper 2000). Unfortunately, the evaluation is very limited and only for very few systems. The most complete overview of technologies to date is given by Smeulders et al (2000). This article describes common problems such as the semantic gap or the sensory gap and gives links to a large number of articles describing the various techniques used in the domain. For an even deeper introduction into the domain, several theses and books are available (Muller 2002, Smith 1997, Del Bimbo 1999, Rahman 2001).
Tang reviewed several medical retrieval systems in the year 1999 (Tang et al 1999). Here, no systematic comparison of the techniques employed and data/evaluation used has been attempted. However, a systematic overview of techniques used, visual features employed, images indexed and medical departments involved have also been explained. It offers future perspectives for image retrieval in the medical domain and is a good starting point for any research work on medical image retrieval system.

### 2.1.1 Content-based Image Retrieval Systems

Although early systems existed already in the beginning of the 1980s (Chang et al 1980), the majority would recall systems such as IBM's QBIC1 (Query by Image Content) as the starting point of content-based image retrieval (Flickner et al 1995, Niblack et al 1993). The commercial QBIC system is definitely the most well-known system. Another commercial system for image (Bach et al 1996) and video (Hampapur et al 1997) retrieval is Virage2 that has well known commercial customers such as CNN.

Most of the available systems would be hard to name or compare. Among them some well known examples include Candid (Kelly et al 1995), Photobook3 (Pentland et al 1996) and Netra (Ma et al 1997) that all use simple color and texture characteristics to describe the image content. Use of higher level information, such as segmented parts of the image for queries, was introduced by the Blobworld4 system (Carson et al 1999, Belongie et al 1998). Pic Hunter (Cox 1996) on the other hand is an image browser that helps the user to find an exact image in the database by showing to the user images on screen that maximize the information gained in each feedback step. A system that is available at free of charge is the GNU Image Finding Tool (GIFT5) (Squire et al 2000a). Some systems are available as demonstration versions on the web such as Viper6, WIPE7 or Compass8.
Most of these systems have a very similar architecture for browsing and archiving/indexing images comprising tools for the extraction of visual features, storage and efficient retrieval of these features, based on distance measurements or similarity calculation and a type of Graphical User Interface (GUI).

2.1.2 Visual Features Used

Visual features were classified in (Eakins et al 2000) into primitive features such as colour or shape, logical features such as identity of objects shown and abstract features such as significance of scenes depicted. Still, all currently available systems use only primitive features unless manual annotation is coupled with the visual features as in (Pfund et al 2002). Even systems using segments and local features such as Blob world (Carson C. et al 1999, Belongie et al 1998) are still far away from identifying objects reliably. No system offers interpretation of images or even medium level concepts as they can easily be captured with text. This loss of information from an image to a representation by features is called the semantic gap (Smeulders et al 2000). The situation is surely not satisfactory and the semantic gap definitely accounts for part of the rejection to use image retrieval applications, but the technology is still valuable when advantages and problems are understood by the users. The more a retrieval application is specialized for a certain, limited domain, the smaller the gap can be made by using domain knowledge. Another gap is the sensory gap that describes the loss between the actual structure and the representation in a (digital) image.

2.1.3 Colour

In stock photography (large, varied databases for being used by artists, advertisers and journalists), colour has been the most effective feature and almost all systems employ colours. Although most of the images are in
the RGB (Red, Green, Blue) color space, this space is only rarely used for indexing and querying as it does not correspond well to the human color perception. It only seems reasonable to be used for images taken under exactly the same conditions each time such as trademark images. Other spaces such as HSV (Hue, Saturation, Value) (Squire et al 2000b, Carson et al 1997, Smith et al 1996) or the CIE Lab (Niblack 1993) and Luv (Sclaro et al 1997) are much better with respect to human perception and are more frequently used. This means that differences in the colour space are similar to the differences in human perception regarding colours.

Many efforts have also been made on creating colour spaces that are optimal with respect to lighting conditions or that are invariant to shades and other influences such as viewing position (Gevers et al 1996, Geusebroek et al 2001). This allows identification of colours even under varying conditions but on the other hand information about the absolute colours is lost. In specialized fields for instance, in the medical domain, absolute colour or gray level features have often very limited expressive power unless the exact reference points exist as in the case of computed tomography images.

2.1.4 Texture

Partly due to the imprecise understanding and definition of what exactly visual texture actually is, texture measures have an even larger variety than color measures. Some of the most common measures for capturing the texture of images are wavelets (Ortega et al 1998, Ze Wang et al 1997) and Gabor filters (Squire et al 2000a, Ma et al 1996, Santini et al 1996) where the Gabor filters do seem to perform better and correspond well to the properties of the human visual cortex for edge detection (Daugman 1990, Daugman 1993). These texture measures try to capture the characteristics of the image or image parts with respect to changes in certain directions and the scale of the changes. This is most useful for regions or images with homogeneous
texture. Again, invariance with respect to rotations of the image, shifts or scale changes can be included into the feature space but information on the texture can get lost in this process (Milanese et al 1999).

Other popular texture descriptors contain features derived from co-occurrence matrices (Shyu et al 1999, Kuo et al 2002), features based on the factors of the Fourier transform (Milanese et al 1999) and the so-called World features (Lu et al 1998).

2.1.5 Local and Global Features

Both, color and texture features can be used on a global image level or on a local level on parts of the image. The easiest way to use regional features is to use blocks of fixed size and location, so-called partitioning of the image for local feature extraction. These blocks do not take into account any semantics of the image itself. When allowing the user to choose image regions (ROI, regions of interest) (Comaniciu et al 1998), to delineate objects in the image (Perry 1997) or when segmenting the image into areas with similar properties (Winter et al 1999), the locally extracted features contain more information about the image objects or underlying structures.

2.1.6 Segmentation and Shape Features

Fully automated segmentation of images into objects itself is an unsolved problem. Even in fairly specialized domains, fully automated segmentation causes many problems and is often not easy to realize. In image retrieval, several systems attempt to perform an automatic segmentation of the images in the collection for feature extraction (Carson et al 1999, Lucchese et al 1999). To have an effective segmentation of images using varied image databases the segmentation process has to be done based on the colour and texture properties of the image regions (Winter et al 1999). Much has also
been written on medical image segmentation with respect to browsing image repositories (Ghebreab 2002, Lapeer et al 2002). After segmentation, the resulting segments can be described by shape features that commonly exist, including those with invariance with respect to shifts, rotations and scaling (Loncaric 1998, Veltkamp et al 2000).

2.2 COMPARISON TECHNIQUES USED

Basically all systems use the assumption of equivalence of an image and its representation in feature space. These systems often use measurement systems such as the easily understandable Euclidean vector space model (Niblack et al 1993, Jain et al 1996) for measuring distances between a query image (represented by its features) and possible results representing all images as feature vectors in an $n$-dimensional vector space. This is done, although metrics shown to not correspond well to human visual perception. Several other distance measures do exist for the vector space model such as the city-block distance, the Mahalanobis distance (Niblack et al 1993) or a simple histogram intersection (Swain et al 1991). Still, the use of high dimensional feature spaces has shown to cause problems and great care needs to be taken with the choice of distance measurement in order to retrieve meaningful results (Aggarwal et al 2001, Hinneburg et al 2000). These problems with a similarity definition in high-dimensional feature spaces are also known as the curse of dimensionality and have also been discussed in the domain of medical imaging (Hanka et al 1996).

Another approach is a probabilistic framework to measure the probability that an image is relevant (Vasconcelos et al 2000a). A relationship between probabilistic image retrieval and vector-space distance measures is given in (Vasconcelos et al 2000b). This paper concluded that the vector space distance measurements described in the literature correspond, in principle, to probabilistic retrieval under certain assumptions of the feature
distributions. Another probabilistic retrieval form is the use of Support Vector Machines (SVMs) (Goh et al 2001) for classification of images into relevant and non-relevant classes.

Various systems use methods that are well known in the text retrieval field and apply them to visual features where the visual features have to correspond roughly to words in text (Squire et al 2000a, Westerveld 2000, Zhu et al 2001). This is based on two principles:

- a feature frequent in an image describes this image well;
- a feature frequent in the collection is a weak indicator to distinguish images from each other.

Several weighting schemes for text retrieval that have also been used in image retrieval are described in (Salton et al 1988). A formal definition of vector-space, probabilistic and boolean models for information retrieval is attempted in (Dominich 2000). A general overview of pattern recognition methods and various comparison techniques are given in a very good review article (Sinha et al 2002). This article describes the feature extraction, selection, features space reduction techniques that are equally important in the image retrieval domain.

2.3 USE OF IMAGE RETRIEVAL IN MEDICAL APPLICATIONS

The number of digitally produced medical images is rising steadily and strongly. In the radiology department of the University Hospital of Geneva (HUG) alone, the number of images produced per day in 2002 was 12,000, and it is still rising. Videos and images produced in cardiology are equally multiplying and endoscopic videos promise to be another very large data source that are planned to be integrated into the PACS. The management
and the access to these large image repositories become increasingly complex. Most access to these systems are based on the patient identification or study characteristics (modality, study description) (Lehmann et al 2003) as it is also defined in the DICOM standard (Revet 1997).

Imaging systems and image archives have often been described as an important economic and clinical factor in the hospital environment (Greenes et al 2000, Kulikowski et al 2002, Vannier et al 2002). Several methods from the computer vision and image processing fields have already been proposed for use in medicine more than ten years ago (Sarvazyan et al 1991, Pun et al 1994). Several radiological teaching files exist (Rosset et al 2002, Binet et al 1995) and radiology reports have also been proposed in a multimedia form in (Maloney et al 1999). Web-interfaces to medical image databases are described in (Frankewitsch et al 2001).

Retrieval systems have often been used for medical images and the medical domain is often cited as one of the principal application domains for content-based access technologies (Beretti et al 2001, Ogiela et al 2001, Orphanoudakis et al 1994) in terms of potential impact. Still, there has rarely been an evaluation of the performance and the description of the clinical use of these systems is even rarer. Two exceptions are the Assert10 system on the classification of high resolution CTs of the lung (Shyu et al 1999, Aisen et al 2003) and the IRMA11 system for the classification of images into anatomical areas, modalities and view points (Keysers et al 2003).

Content-based retrieval has also been proposed several times from the medical community for the inclusion into various applications (Tagare et al 1997, Lowe et al 1998, Bidgood et al 1999), often without any implementation. For a real medical application of content-based retrieval methods and the integration of these tools into medical practice a very close cooperation between the two fields viz., medical and image processing is
necessary for a longer period of time. Simply an exchange of data or a list of the necessary functionality among them is not sufficient.

2.4 TECHNIQUES USED IN MEDICAL IMAGE RETRIEVAL

This section describes the various techniques that are currently used or that have been proposed for the use in medical image retrieval applications. Many of the techniques are similar to those used for general content-based retrieval but have not yet been used in medical applications. A special focus is put on the data sets that are used to evaluate the image retrieval systems and on the measurements used for evaluation. Unfortunately, the performance evaluation of systems is currently remaining neglected.

Machine learning in medical applications also gets increasingly more important and it is essential to research the various possibilities. Specialized workshops exist for this purpose (Magoulas et al 2001).

2.4.1 Features Used

This section describes the (visual) features that are used in the various applications. The section text is added to discuss whether this should be named content-based retrieval or not. As the formulation of similarity queries without text can be quite a problem, another subsection is added to describe the various possibilities to formulate queries without text.

2.4.2 Query Formulation

The query formulation using exclusively visual features can be a big problem. Most systems in CBIR use the Query by Example (QBE) paradigm which needs an appropriate starting image for querying. This problem of a sometimes missing starting image is known as the page zero problems.
If text is attached to the images, which is normally the case in medical applications, then the text can be used as a starting point and once visually relevant images have been found, further queries can be entirely visual (Orphanoudakis et al 1996) to find visually similar cases not found by text or to sort the found cases by their visual similarity. In the medical decision-making process, there are often images produced and available for the current case. The starting point does not need to be further defined but the images of the case can be used directly (El-Kwae et al 2000). In connection with the segmentation of the images the user can also restrict the query to a certain Region of Interest (ROI) in the image (El-Kwae et al 2000), which can lead to much more specific queries than if using an image in its entirety.

The use of human sketches has already been proposed in generic image retrieval (Egenhofer 1996) and it has also been proposed for the use in medical applications (Antani et al 2002, Le Bozec et al 2000, Ikeda et al 2000). Considering the difficulty in exact drawing and the need for some artistic skills and time, this method could be applied for a very small subset of queries, such as tumor shapes or spine x-rays, where outlines are possible directly in the image. For general image retrieval, sketches are too time-consuming and the retrieved results are often not exact enough.

2.4.3 Text

Many systems propose use of text from the patient records (Le Bozec et al 2000) or studies (El-Kwae et al 2000) to search by content. Others define a context-free grammar, a standardized vocabulary for image description (Jaulent et al 2000) or an image definition language (Traina Jr. et al 1997) for the querying of images in image repositories. The combination of textual with visual features or content and context of the images does have the most potential to lead to good results (Antani et al 2002). One can also be
used to control the quality of the other or to obtain a better recall of the retrieval results.

Besides the free text that is frequently used for retrieval, medical patient records also contain very valuable structured information such as age, sex and profession of the patient. This information is just as important as free text to put the images into a context.

2.4.4 Visual Features

Unfortunately, most articles that propose content-based queries do not explain in detail which visual features have been used or are planned to be used. Sometimes, only a very vague description such as general texture and color or gray level features are given as in (Mojsilovis et al 2000, Chbeir et al 2000).

Basically all systems that do give details use color and gray level features, mostly in the form of a histogram (Tang et al 2000, Kwak et al 2002). Local and global gray level features are used in (Muller et al 2003). (Keysers et al 2003, Guld et al 2001) used statistical distributions of gray levels for the classification of images. As many of the images in the medical domain do not contain colors or are taken under controlled conditions, the color properties are not at all in the center of research and the same holds good for invariants to lighting conditions. These would get changed when using photographs such as in dermatology. Pathologic images will need to be normalized in some way as different staining methods can produce different colors (Wurflinger et al 2003). Within radiology, the normalization of gray levels among different modalities or even for the same modality can cause problems when there is no exact reference point as is for the density of the CT, for example. Features derived from co-occurrence matrices are also frequently used (Beretti et al 2001, Kwak et al 2002). In mammography,
denseness is used for finding small nodules (Baeg et al 2002). It would be interesting to have a comparison of several texture descriptors. Many of them model the same information and will most likely to deliver very similar results.

2.5 COMPARISON METHODS AND FEATURE SPACE REDUCTIONS

Most systems do not give many details on the distance measurements or comparison methods used which most likely implies an Euclidian vector space model using either a simple Euclidean distance (L2) or something close to it such as city block distance or L1. To work efficiently with these distances even in large databases, the dimensionality is often reduced. This is done with methods such as Principal Component Analysis (PCA) (Sinha et al 2002, Bucci et al 1996) or Minimum Description Length (MDL) (Brodley et al 1999) that try to reduce the dimensionality while staying as discriminative as possible. In principle, redundant information is removed but this can also remove small but important changes from the feature space.

On the other hand, statistical methods are used for the comparison of features that can be trained with existing data and that can then be used on new, incoming cases. These can be neural networks for the classification of mammography images (Baeg et al 2002) or on images extremely reduced in size (18×12 pixels) in (Ikeda et al 2000). Other statistical approaches use Bayesian networks (Liu et al 1997) or Hidden Markov Models (HMMs) (Beretti et al 2001). In (Khan et al 1996), an associative computing approach is proposed for retrieval assuming that a query is performed with a local part of the images.
The preceding subsections already showed the large variability in techniques that are used for the retrieval of images. Still, several very successful techniques from the image retrieval domain have not yet been used for medical images. The entire discussion on relevance feedback that first improved the performance of text retrieval systems and then, 30 years later, of image retrieval systems has not at all been discussed for the medical domain. A few articles mentioned it but without any details on use and performance. Often the argument for omitting relevance feedback is that medical doctors do not have the time to look at cases and judge them. If the systems are interactive (response times below 1 second), (Nielsen 1993) this should not be a reason as an expert can mark few images as positive and negative relevance feedback within a minute or less. Also the prospect of long-term learning from this marking of images should motivate people to use it. Long-term learning has shown to be an extremely effective tool for system improvements.

Another domain not discussed at all for medical images are the user interfaces. Sometimes web-based interfaces are proposed (Muller et al 2003, Lehmann et al 2003) but no comparison of interfaces is reported and no real usability studies have been published to the author’s knowledge so far. As there are several creative solutions in image retrieval it will be interesting to study the effects of interfaces, ergonomics and usability issues on the acceptance and use of the technology in clinical practice.

Performance comparisons for different feature sets have also never been performed and are important to identify well-performing visual features and the applications that they can successfully be used for. This helped us a great deal to start this thesis with different features in the domain and also to optimize existing systems.