Contents

21

Content-Based Medical Image Retrieval

Henning Müller and Thomas M. Deserno

Abstract. This chapter details the necessity for alternative access methods to the currently mainly text-based methods in medical information retrieval. This need is partly due to the large amount of visual data produced, the increasing variety of medical imaging data and changing user patterns. The stored visual data contain large amounts of unused information that, if well exploited, can help diagnosis, teaching and research. The chapter briefly reviews the history of image retrieval and its general methods before technologies that have been developed in the medical domain are discussed. We also discuss evaluation of medical content-based image retrieval (CBIR) systems and conclude with pointing out their strengths, gaps, and further developments. As examples, the MedGIFT project and the Image Retrieval in Medical Applications (IRMA) framework are presented.

21.1 Introduction

Content-Based Visual Information Retrieval (CBVIR) or Content-Based Image Retrieval (CBIR) has been one on the most vivid research areas in the field of computer vision over the past almost 20 years. The availability of large and steadily growing amounts of visual and multimedia data, and the development of the Internet underline the need to create access methods that offer more than simple text-based queries or requests based on matching exact database fields. Many programs and tools have been developed to formulate and execute queries based on the visual or audio content and to help browsing large multimedia repositories. Still, no general breakthrough has been achieved with respect to large varied databases with documents of differing sorts and with varying characteristics. Answers to many questions with respect to speed, semantic descriptors or objective image interpretations are still open and wait for future systems to fill the void [1].

In the medical field, images, and especially digital images, are produced in ever-increasing quantities and used for diagnosis and therapy. The Radiology Department of the University Hospitals of Geneva alone produced more than 114,000 images a day in 2009, risen form 12,000 in 2002. Large hospital groups such as Kaiser Permanente that manage several hospitals had by early 2009 even 700 TB of data stored in the institutional archives and very large hospitals such as the University hospital of Vienna currently produces over 100 GB of image data per day.

With Digital Imaging and Communications in Medicine (DICOM), a standard for image communication has been set and patient information can be stored with the actual image(s), although still a few problems prevail with respect to the standardization. In several articles, content-based access to medical images for supporting clinical decision-making has been proposed [1, 2]. Still, only very few systems are usable and used in clinical practice as most often development takes place in computer science departments totally disconnected from clinical practice.

21.1.1 Motivation and History

Image retrieval has been an extremely active research with first review articles on access methods in image databases appearing already in the early 1980s [3]. The following review articles explain the state-of-the-art and contain references to a large number of systems and descriptions of the technologies implemented [4–7]. The most complete overview of technologies to date is given by Smeulders et al. in [8]. 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. In a more recent article, the developments over the past 5-10 years are described [9].

Although early systems existed already in the beginning of the 1980s [10], the majority would recall systems such as IBM's Query by Image Content (QBIC)[1] as the start of content-based image retrieval [11].

Most of the available systems are, however from academia. It would be hard to name or compare them all but some well-known examples include Photobook [12] and Netra [13] that all use simple color and texture characteristics to describe the image content. Using higher level information, such as segmented parts of the image for queries, was introduced by the Blobworld[2] system [14, 15]. PicHunter [16] 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 gain in each feedback step. A system that is available free of charge is the GNU Image Finding Tool (GIFT)[3] [17].

21.1.2 Query-by-Example(s) Paradigm

One of the biggest problems in CBIR is the formulation of queries without text. Everyone is used to formulate queries with text (as 90 % of Internet

[1] `http://wwwqbic.almaden.ibm.com/`
[2] `http://elib.cs.berkeley.edu/photos/blobworld/`
[3] `http://www.gnu.org/software/gift/`

users are using Google) and explain one's information needs but with visual elements this is far from trivial. Drawing small designs is one possibility requiring artistic skills and being unsuitable for the majority of users. Using image examples to search with Query by Image Example (QBE) is currently the most common way to search for similar images being used by most image retrieval systems. Thus, a system can search for visually similar images with one or several example image(s). The problem remaining is to find a suitable example image, which is not always obvious ("page zero problem") [18].

In the medical domain images are usually one of the first examinations performed on patients, and thus query examples are available. Once the user has received a results set of images or cases similar to a given example image or case, systems most often offer the possibility to mark images/cases as relevant and irrelevant and thus refine the search through what is called ("relevance feedback") [19].

21.2 General Image Retrieval

General image retrieval started with the main concepts already in 1980 [3]. Still, the real research did not start before the late 1980s, when several systems using simple visual features became available [11].

21.2.1 Classification vs. Retrieval

One of the first and basic questions in image retrieval is whether it is rather an information retrieval task or a classification task. While there are many similarities between them, there are two principle differences [20]:

- *classification* tasks have a limited number of classes of topics/items and training data for each of the classes that allow training of class-specific parameters;
- *retrieval* tasks have no fixed classes of items/objects in the database and usually no training data available; documents can be relevant for a particular retrieval task or information need, with relevance being potentially user-dependent.

In general, the techniques according to the classification paradigm follow the general machine learning literature and its approaches, whereas the (information) retrieval approaches follow techniques from general information retrieval.

In the end, when used for CBIR, both represent images by visual features and then find similar images using a distance measure, showing the most similar images to the user ordered by their visual similarity.

21.2.2 System Setup, Components, Computation, Grid Networks

Most of these systems have a very similar architecture for browsing and archiving/indexing images comprising tools for the extraction of visual features, for the storage and efficient retrieval of these features, for distance measures or similarity calculation and a type of Graphical User Interface (GUI). This general system setup is shown in Fig. 21.1.

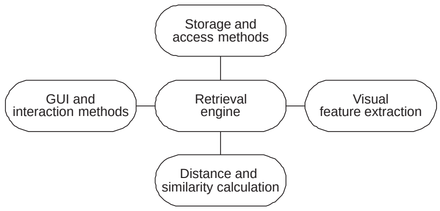

Fig. 21.1. Retrieval system architecture. Overview of the main components that most image retrieval systems are constituted of.

Computational efficiency is another often regarded question. Particularly the visual analysis can take an enormous time for large databases and as the challenge is to scale to millions of images, tools such as Grid networks and parallel processing have been used for the off-line feature processing. This is mainly used for the off-line step of representing images by features, whereas for the query processing efficient indexing structures are used for quick response times $t_r < 1\,\mathrm{s}$.

21.2.3 Features and Signatures

Visual features were classified into *primitive* features such as color or shape, *logical* features such as identity of objects shown and *abstract* features such as significance of depicted scenes [6]. However, basically all currently available systems only use primitive features such as:

- *Color*: In stock photography (large, varied databases for being used by artists, advertisers and journalists), color has been the most effective feature. The red, green, blue (RGB) color space is only rarely used as it does not correspond well to the human color perception. Other spaces such as Hue-Saturation-Value (HSV) or the CIE Lab and Luv spaces perform better because distances in the color space are similar to the differences between colors that humans perceive. Much effort has also been spent on creating color spaces that are optimal with respect to lighting conditions or that are invariant to shades and other influences such as viewing position [21].
- *Texture*: Texture measures try to capture the characteristics of the image with respect to changes in certain directions and the scale of the changes. This is most useful for images with homogeneous texture. Some of the most common measures for capturing the texture of images are wavelets and Gabor filters. Invariances with respect to rotation, shift or scale can be included into the feature space but information on the texture may get lost in this process [22]. Other popular texture descriptors contain features

derived from co-occurrence matrices [23, 24], the Fourier transform [22], and the so-called *Wold* features [25].
- *Local color and texture*: Both, color and texture features can be used also on a local or regional level, i.e. on parts of the image. To use blocks of fixed size, so-called *partitioning*, is the easiest way employing regional features [26]. These blocks do not take into account any semantics of the image itself. When allowing the user to choose a Region of Interest (ROI) [27], or when segmenting the image into areas with similar properties [28], local features capture more information about relevant image structures.
- *Shape*: Fully automated segmentation of images into objects itself is an unsolved problem. Even in fairly specialized domains, automated segmentation causes many problems. In image retrieval, several systems attempt to perform an automatic segmentation for feature extraction [29]. The segmentation process should be based on color *and* texture properties of the image regions [28]. The segments can then be described by shape features, usually being invariant to shift, rotation and scaling [30]. Medical image segmentation with respect to browsing image repositories is frequently addressed in the literature as well [31].
- *Salient points*: Salient point-based features have in recent years had best performances in most of the image retrieval and object classification tasks [32]. The idea is to find representative points (or points that attract the attention) in the images and then analyze the relationships of the points. This permits to extract features that possess several invariants such as invariance to shifts, rotations, scale and even view-point. A large number of such techniques exist for detecting the points and then for extracting features from the salient points.
- *Patches and visual words*: Patches and visual words are closely linked to salient point-based features. as the patches and/or visual words are most often extracted from regions in the images that were identified to contain changes or high gradients and then local features are extracted in these regions. It is also possible to put a regular grid on the image and then extract patches around the points of the grid to well represent the entire image. The term visual words stems from the fact that the features extracted around the selected points are often clustered into a limited number of homogeneous characteristics that can have distributions similar to the distribution of words in text allowing to use techniques well known from text retrieval [33].

All of these features have their benefits and domains where they clearly work well, but all these features are low-level visual features and might not correspond to semantic categories. For this reason, text, whenever available should be used for the retrieval of images as well, as semantic information is conveyed very easily. All benchmarks show that text has a much superior performance compared to visual characteristics, but can well be complemented by visual retrieval.

21.2.4 Distance and Similarity Measures

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 [11] 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 have been shown to not correspond well to human visual perception [34]. Several other distance measures do exist for the vector space model such as the city-block distance, the Mahalanobis distance [11] or a simple histogram intersection [35]. 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 to be chosen in order to retrieve meaningful results [36, 37]. These problems with a similarity definition in high-dimensional feature spaces is also known as the "curse of dimensionality" and has also been discussed in the domain of medical imaging [38].

Another approach is a *probabilistic framework* to measure the probability that an image is relevant [39]. Another probabilistic retrieval form is the use of a Support Vector Machine (SVM) [40, 41] for classification of images into classes for relevant and non-relevant. In most visual classification tasks SVMs reach in general the best performance.

Various systems use methods that are well known from the text retrieval field and apply them to visual features where the visual features have to correspond roughly to words in text [26, 42]. This is based on the 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 [43]. A general overview of pattern recognition methods and various comparison techniques is given in [44].

21.3 Medical Image Retrieval

The number of digitally produced medical images has rising strongly, mainly due to large tomographic series. 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 many Picture Archiving and Communication Systems (PACS). The management and the access to these large image repositories become increasingly complex. Most accesses to these systems are based on the patient identification or study characteristics (modality, study description) [45].

Imaging systems and image archives have often been described as an important economic and clinical factor in the hospital environment [46, 47]. Several methods from the computer vision and image processing fields have already been proposed for the use in medicine more than ten years ago [48]. Several radiological teaching files exist [49] and radiology reports have also been proposed in a multimedia form in [50].

21.3.1 Application Fields

Content-based retrieval has also been proposed several times from the medical community for the inclusion into various applications [2,51], often without any implementation. Figure 21.2 shows the general system architecture.

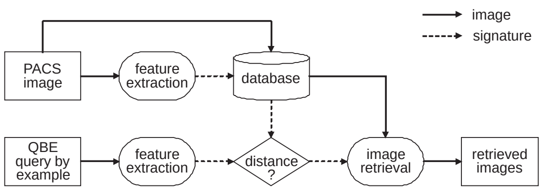

Fig. 21.2. Medical CBIR system architecture. All images in the PACS archive and the QBE image are describes by a signature. Comparing signatures instead of images allows fast CBIR response.

Almost all sorts of images have already been used for image retrieval at one point or another. The first separation is on whether systems use a large and varied set of images [52] or work on a very focused domain as diagnosis aid [53].

A typical application domain for CBIR-based image management is case-based reasoning and evidence-based medicine, in particular in fields where diagnosis is regarded as hard and where purely visual properties play an important role, such as mammography [54] or the diagnosis of interstitial lung diseases [55, 56]. CBIR-based eLearning has also been discussed [57].

21.3.2 Types of Images

The medical domain yields an extremely large amount of varying images, and only very few have so far been exploited fully for visual similarity retrieval. When thinking of medical images, clearly radiographs and maybe computed tomography come instantly to mind but there is much more than this usually gray scale set of images.

Here is a list of some of the types of visual data that are available in hospitals and often stored in the PACS:

- 1D signals: EEG, ECG;
- 2D gray scale images: x-ray radiography;
- 2D color images: pathology slides microscopy, photography, dermatology;

- gray scale video: ultra-sonography;
- color video: sleeping laboratory;
- pseudo-3D (slices): CT, MRI, PET, SPECT;
- 3D models: reconstructions of tomographic images;
- 4D data: temporal series of tomographic images such as CT images of a beating heart;
- nD data: Multi-modal combinations of images such as from combined PET/CT, PET/MRI scanners.

It becomes quickly clear that medical imaging is much more varied then the images of the general CBIR domains, such as photographs in the Internet.

21.3.3 Image Preprocessing

Image pretreatment is most often used to harmonize the content in a database and thus make feature extraction from the images based on the same grounds. Such preprocessing can be the normalization of gray levels or colors in images.

Another application of pretreatment in the medical domain is the background removal from images and automatic detection of the field of view [58] to concentrate the search on the important objects. Although medical images are taken under relatively controlled conditions, there is a fairly large variety remaining particularly in collections of scanned images.

Some typical images from our database are shown in Fig. 21.3 (top row). The removal is mainly done through a removal of specific structures followed by a low pass filter (median) and then by thresholding and a removal of small unconnected objects. After the object extraction phase, most of the background is removed but only a few images had part of the main object removed (Fig. 21.3, bottom row).

21.3.4 Visual and Non-Visual Image Features

Most of the visual features used in medical images are based on those existing for non-medical imaging as well [59]. For radiographs, there is clearly a need to highlight gray level values instead of the color values in non medical image, which can make the search harder. On the other hand most of the medical images are taken under fairly standardized conditions, requiring fewer invariances and even making direct comparisons of downscaled versions of the images possible.

In contrast to non-medical image archives, all medical images do have meta-data attached to them as the images are part of a clinical record, that consists of large amounts of structured data and of free text such as laboratory results, anamnesis and release letter. Without this meta information, interpretation of medical cases is impossible. No radiologist would read an image without a minimum of meta data on the patient (e.g., age, sex) and a basic anamnesis as many of the clinical parameters do have a strong influence on the visual characteristics of the images. For instance, old patients have less dense bones, and the healthy lung of a smoker is much different from the healthy lung of a non-smoker.

One of the largest problems is how to combine structured/free text data with visual features. Several fusion approaches have been proposed in [56]. Most often, late fusion is considered best as there are potentially many features and there can be negative interactions between certain of the clinical data and certain visual features. It is also clear that the data quality in patient records is often far from optimal and even in an anamnesis not all parameters are asked systematically, leaving often incompleteness, for example, where it is not clear if the patient was a smoker or not. Such incomplete data needs to be taken into account for classification or retrieval approaches [59].

21.3.5 Database Architectures

Many tools and techniques have been used to allow for a quick access in large collections of images, similar to access models in general database ar-

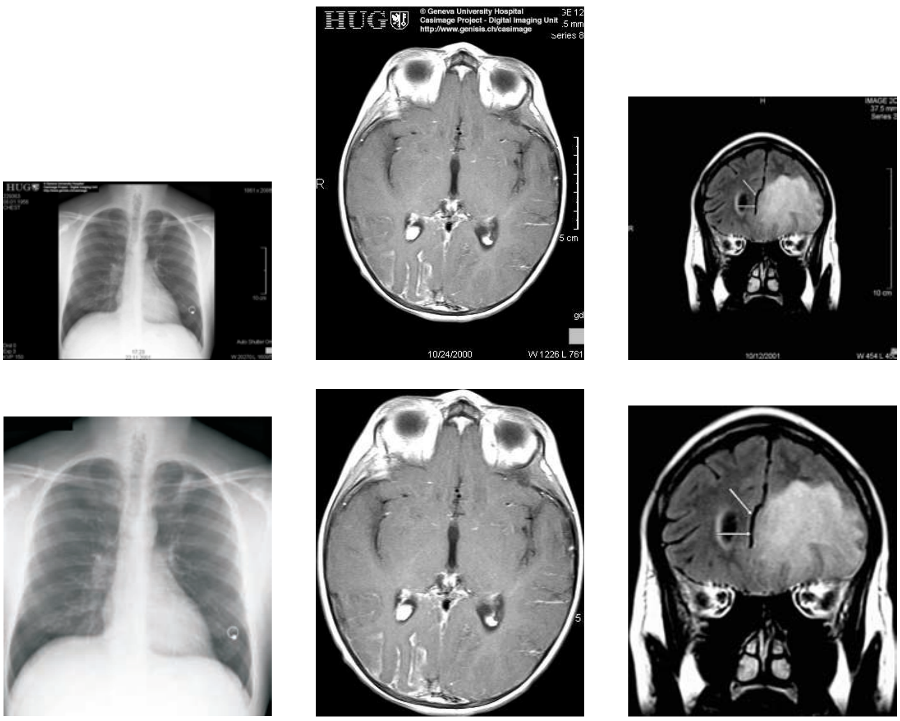

Fig. 21.3. Image Pretreatment. Images before (top row) and after (bottom row) the removal of logos and text.

chitectures. Frequently, the goal is to have rather a long off-line phase of data pretreatment followed by a rather quick query response. Techniques from text retrieval have shown to allow for extremely quick response times in sparsely populated spaces and are frequently used.

Parallel access to databases and grid networks are frequently used for the off-line phase, so the most computationally heavy phase. For on-line processing this is often too slow, though, as often there is an overhead in grid networks, for example, for the job submission and load balancing part.

21.3.6 User Interfaces and Interaction

Most of the current user interfaces follow the QBE paradigm and allow to upload images to start with, or have a random function to browse images in the database to find a starting point. Most interfaces show a ranked list of results images ordered by similarity. A clear distinction needs to be made for how visual and how textual queries can be formulated. Both together form the most powerful framework [60].

Another important aspect of the user interface is the possibility to obtain more information about the users information need by marking images as positive and/or negative feedback. Many techniques exist for calculating similarity between several positive and negative input images, from combining all features for a joint pseudo-image to performing separate queries with each image and then combining the results.

21.3.7 Interfacing with Clinical Information Systems

The use of content-based techniques in a PACS environment has been proposed several times [61]. PACS are the main software components to store and access the large amount of visual data used in medical departments [62]. Often, several layer architectures exist for quick short-term access and slow long-term storage, but these are steadily replaced by fully hard disk oriented solutions. The general schema of a PACS system within the hospital is shown in Fig. 21.4. The Integrating the Healthcare Enterprise (IHE) initiative is aiming at data integration in healthcare including all system components.

An indexing of the entire PACS causes problems with respect to the sheer amount of data that needs to be processed to allow efficient access by content to all the images. This issue of the amount of data that needs to be indexed is not discussed in any of the articles. Qi & Snyder have proposed to use CBIR techniques in a PACS as a search method but no implementation details are given [63]. Bueno et al. extend a database management system for integrating content-based queries based on simple visual features into PACS [64]. A coupling of image classification with PACS is given in [45]. Here, it is possible to search for certain anatomic regions, modalities or views of an image. A simple interface for coupling the PACS and the image retrieval system is also proposed. The identification is based on the DICOM Unique Identifier (UID)

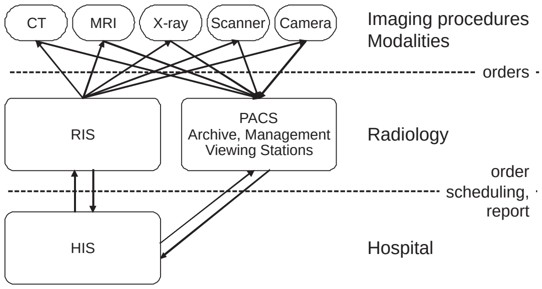

Fig. 21.4. System interconnection. The Picture Archiving and Communication System (PACS) is connected with imaging modalities such as Computed Tomography (CT) or Magnetic Resonance Imaging (MRI), the Radiology (RIS) and the Hospital Information System (HIS).

of the images. An IHE compliant procedure calling external CBIR application as well as returning the CBIR results into the PACS are described [65, 66].

21.4 Evaluation

Whereas early evaluation in image retrieval was only base on small databases showing a few example images, evaluation in text retrieval has always been a very experimental domain. In CBIR, a first real standardization was achieved place with the ImageCLEF[4] medical image retrieval task that started in 2004 as has been organized every year since, including and classification task and a retrieval task based on a data set of the Image Retrieval in Medical Applications (IRMA)[5] group.

21.4.1 Available Databases

Medical image databases have increasingly become available for researcher in the past five years. Some of the prominent examples is the Lung Image Database Consortium (LIDC) data, the IRMA database with many different classes and an increasing number of images and the images of the ImageCLEF competition taken first from medical teaching files and then from the scientific medical literature.

Nowadays, the National Institutes of Health (NIH) and the National Cancer Institute (NCI) require funded research to make their data available, and several databases indeed have become available for the public.

21.4.2 Tasks and User Models

When evaluation image retrieval it is clear that a clear usage model and information need have to be defined. A few research groups have actually conducted

[4] `http://www.imageclef.org/`
[5] `http://irma-project.org`

surveys on the use of images for journalists [67] and in other domains such as libraries or cultural heritage institutions [4].

For ImageCLEF 2005 the topic development was based on two surveys performed in Portland, OR and in Geneva [68,69]. In total, around 40 medical professionals were surveyed on their image use and search behavior to learn more on how they use images and how they would like to search for them. It became clear that depending on the role of the person (clinician, lecturer, researcher) the information needs are significantly different, so each person who had more than one role had to respond to the questions for all roles. Librarians and students were also included into the survey. Most frequently, people said that they would like to be able to search for pathology and then came modality and anatomic region. People said to use Web search engine to search for interesting images for lectures. People were concerned about image quality in this case. Based on these surveys, topics for ImageCLEFmed were developed along the following axes:

- anatomic region shown in the image;
- image modality (e.g., x-ray, CT, MRI, microscopy);
- pathology or disease shown in the image;
- abnormal visual observation (e.g., enlarged heart).

It was tried that topics covered at least two of these axes if possible. A visual query topic is shown in Fig. 21.5, and a query topic requiring more than purely visual features is shown in Fig. 21.6. As ImageCLEF is on multilingual information retrieval and as the collection is in three languages, the topics were also developed in these three languages.

Fig. 21.5. Visual query. An example of a query (topic) of ImageCLEF 2005 that is at least partly solvable visually, using the image and the text as query. Still, use of annotation can augment retrieval quality. The query text is presented in three languages, *English*: "Show me chest CT images with emphysema"; *German*: "Zeige mir Lungen CTs mit einem Emphysem"; *French*: "Montre-moi des CTs pulmonaires avec un emphysème".

21.4.3 Ground Truth and Gold Standards

One of the most important aspects of evaluation is that there is a clear idea of what a good or perfect query result would be like. In the case of the IRMA

Fig. 21.6. Semantic query. A query of ImageCLEF 2005; *English*: "Show me all x-ray images showing fractures"; *German*: "Zeige mir Röntgenbilder mit Brüchen"; *French*: "Montres-moi des radiographies avec des fractures", which requires more than only visual retrieval. Visual features, however, can deliver hints to good results.

collection, this ground truth (or gold standard) is given by the IRMA code that is attributed to each image by a clinician [70]. Its mono-hierarchical multi-axial architecture allows unambiguous ground truth labeling. Therefore, depending on the data sets, classes can be generated using the entire hierarchy or a partial hierarchy. Image classification systems can then be evaluated by comparing them to the correct class labels.

For image retrieval evaluation as in the ImageCLEFmed retrieval benchmark is slightly different as no fixed classes exist. Based on well-defined information such as those in Fig. 21.6 experts can judge whether an image is relevant to this query or not. In images three categories were used, relevant, irrelevant, or indeterminable. Based on the judgments of clinicians on such relevance, several retrieval systems can well be compared

Performance measures for the evaluation of information retrieval in general and image retrieval in particular have created a large amount of discussion for many years. Where as in image classification the choice is smaller (correctly classified, incorrectly classified), there are many measures existing for retrieval tasks.

21.4.4 Benchmarks, Events, and their Limitations

Information retrieval benchmarks have been established in the 1960s with the Cranfield tests. Since 1991, the Text Retrieval Conference (TREC) has created a strong testbed for information retrieval evaluation. For several years, TREC contained a biomedical retrieval called TRECgenomics.

Cross Language Evaluation Forum (CLEF) started within TREC in 1997 and has been independent since 2000. With ImageCLEF that started in 2003, a new medical task was introduced as well, promoting the search for medical images with textual and visual means combined. ¿From a small database of 8,000 images in 2004 the data sets and tasks have grown larger and more complicated every year. Also regarding the IRMA database and the image classification task, the complexity over four years was increased every year.

21.5 Examples for Medical CBIR Systems

This section describes three example projects for medical content-based image retrieval.

21.5.1 MedGIFT

Initially, the Medical GIFT (MedGIFT)[6] project was based on the GNU Image Finding Tool (GIFT), which resulted from the Viper[7] project at the University of Geneva [26]. The visual features used are meant for color photography and include a simple color histogram as well as color blocks in various areas of the images and at several scales. To separate the actual query engine from a user interface, the Multimedia Retrieval Markup Language (MRML)[8] was developed. This query language is based on direct communication of search engine and interface via sockets and eases a variety of applications such as meta-search engines and also the integration of a retrieval tool into a variety of environments and applications.

After a while, however, it became clear that new techniques were necessary in the medical domain, and the build components were grouped around five axes:

- data access, ontologies, data annotation;
- techniques for retrieval and efficient structures to use them on large data sets;
- applications in the medical field such as lung image retrieval, fracture retrieval;
- inclusion of higher dimensional dat sources into the retrieval process such as the use of 3D and 4D data;
- evaluation, mainly with the ImageCLEF benchmark describes in Section 21.4.

Fig. 21.7 shows a typical web interface after a query was executed. The query results are displayed ordered by their visual similarity to the query, with a similarity score shown underneath the images as well as the diagnosis. A click on the image links with the case database system and allows to access the full-size images.

[6] http://www.sim.hcuge.ch/medgift/
[7] http://viper.unige.ch/
[8] http://www.mrml.net/

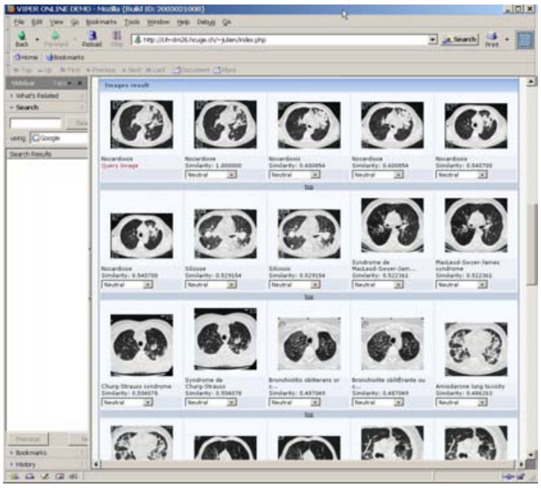

Fig. 21.7. MedGIFT user interface. A screen shot of a typical web interface for medical image retrieval system allowing QBE with the diagnosis underneath the image.

In the context of heading towards indexing of higher-dimensional images an interface for browsing 3D repositions was developed [71] and is shown in Fig. 21.8.

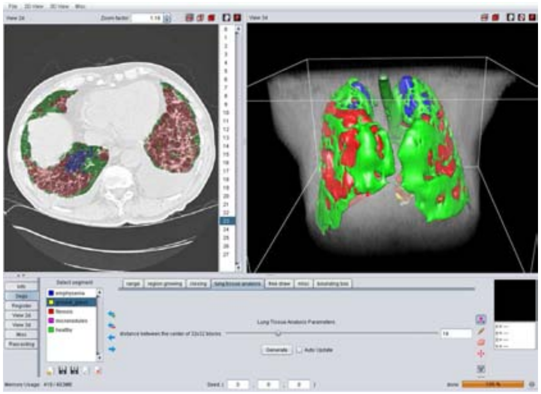

Fig. 21.8. CBIR user interface supporting 3D data. An interface that allows searching in 3D databases by visual content and then can visualized the images with abnormal regions marked in various colors.

21.5.2 IRMA

In Section 21.4, we have already introduced the Image Retrieval in Medical Applications (IRMA) framework. This research-driven project is active for almost ten years, combining inter-disciplinary expertise from diagnostic radiology, computer science, and medical informatics.

IRMA aims at developing and implementing high-level methods for CBIR including prototypical application (e.g., [41,72,73]) to medico-diagnostic tasks on a radiological image archive. They want to perform semantic and formalized queries on the medical image database which includes intra- and inter-individual variance and diseases.

Based on a (i) central database that hold images, features, and the processing methods, (ii) a scheduler that provides distributed processing, (iii) a communicator that is used to interconnect CBIR with Radiology Information System (RIS) and Picture Archiving and Communication System (PACS), (iv) web-based user interfaces are provided for applications citeLGT2003. Three levels of image content similarity are modeled (Fig. 21.9):

- *global features* are linked to the entire images and used to automatically classify an image according to the anatomy, biosystem, creation, and direction (registered data layer) [70],
- *local features* are linked to prominent image regions and used for object recognition (feature layer), and
- *structural features* are linked to spatial or temporal relations between the objects and used for high-level image interpretation (object layer).

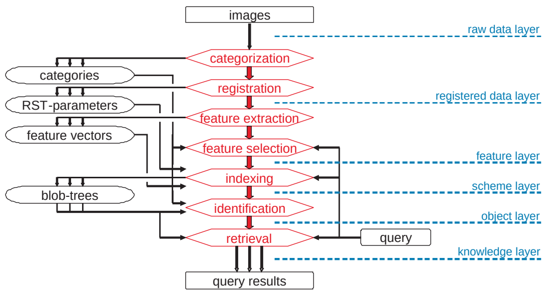

Fig. 21.9. IRMA processing pipeline and levels of content abstraction [45]. All images are categorized and registered to a category-specific prototype. Feature extraction and selection are separated. Constellation of identified objects (scene analysis) are modeled on the highest level.

A pipeline of image processing is suggested (Fig. 21.9). Iterative region merging is used to build up a Hierarchical Attributed Region Adjacency Graph (HARAG), the data structure that is used to represent images, Objects of Interest (OOIs), and object constellations (scene analysis). Hence, image retrieval is transformed to graph matching. Object comparison operates on the HARAG nodes, while scenes are modeled by graph to sub-graph comparison.

Extended query refinement is provided to the user and allows for undo and redo commands and logical combinations of individual query responses [74]. Figure 21.10 visualized the interaction loops that are all encapsulated within one browser window. Parameter modules are used to transfer the input and parameters from the user to the system (e.g., QBE), and the output modules are used to display the query result (Fig. 21.10, green). Query refinement is supported by the orange loop, and yellow indicates undo, and redo options. The outer loop (Fig. 21.10, blue) allows combining individual queries by AND

and OR. Here, the user can seek images having a certain characteristic in one local area and another elsewhere.

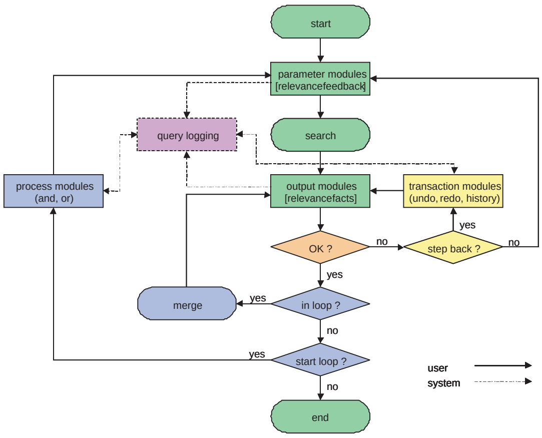

Fig. 21.10. IRMA extended query refinement [74]. Four nested loops are integrated within one browser interface. Green: simple QBE; Orange: query refinement; Yellow: undo and redo; Blue: AND and OR.

A typical IRMA user interface is shown in Figure 21.11. Here, a spine x-ray databased is searched by shape and shape similarity [75]. The slider bars below the images allow the user to evaluate the retrieval result (query refinement).

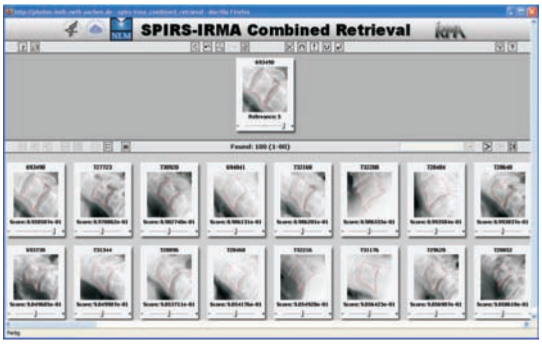

Fig. 21.11. IRMA user interface. A typical IRMA web interface supporting QBE, relevance feedback, and extended query refinement. Here, a shape retrieval interface in collaboration with the National Library of Medicine (NLM), National Institutes of Health (NIH), USA is shown.

Currently, the IRMA group works on integration of CBIR into the clinical workflow. Figure 21.12 shows the dataflow for CBIR-assisted pre-fetching of previous radiographs supporting the radiologist in reading the scheduled exam. Both, Health Level Seven (HL7) and Digital Imaging and Communications in Medicine (DICOM) interfaces are provided by the IRMA communicator module. The dataflow

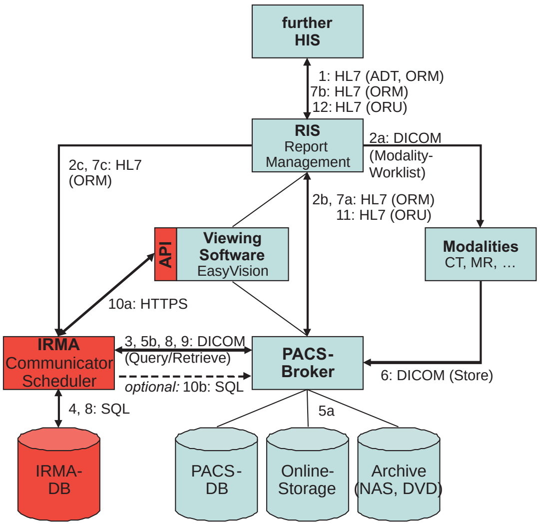

Fig. 21.12. IRMA integration with HIS and PACS [76]. The communication steps are performed (i) at time of examination scheduling (steps 1 to 4); (ii) in the night before the exam (step 4); and (iii) on the day of the examination (steps 5 to 12). The additional communication steps that have been added to the communication because of CBIR integration are: 2c, 3, 6c, 7, 8, 9, 10a. To support CBIR-based hanging protocols, steps 10b and 11b are required additionally.

21.6 Discussion and Conclusions

Medical images have often been used for retrieval systems and the medical domain is often cited as one of the principal application domains for content-based access technologies [77, 78] in terms of potential impact. Still, there has rarely been an evaluation of the performance and the description of the clinical use of systems is even rarer. Two exceptions are the Assert system on the classification of high resolution CTs of the lung [53] and the IRMA system for the classification of images into anatomical areas, modalities and view points [52].

Still, 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 is necessary for a longer period of time and not simply an exchange of data or a list of the necessary functionality.

21.6.1 Strengths and Weaknesses of Current Systems

It was clearly described in this chapter that image retrieval has gone a long way from purely theoretical laboratory style developments, where single images were classified into a small number of classes without any clinical application, towards tools that combine visual and clinical data to really aid diagnosis and deliver valuable information to the clinicians. Tools have have shown to improve diagnosis in real settings when properly applied [79]. With ImageCLEF there is also a benchmark to compare techniques for visual classification as well as for multi-modal medical information retrieval combining text and image data. Such benchmarks are necessary to proof the performance of techniques and entire systems

Still, there is currently a total lack of system that are used in clinical practice and in close collaboration with clinicians.

21.6.2 Gaps of Medical CBIR Systems

In [59, 80], several technical gaps in medical image retrieval have been identified (Fig. 21.13). However, there are several other levels of gaps that need to be mentioned in this context. Legal constraints currently limit the application domain as the secondary use of medical data is ruled by nationally different laws that are not always easy to follow. In general, informed consent is required even if data is anonymized. This limits the amount of data potentially accessible and thus also the usefulness of the approach. Tools as the one described in [81] to access research data in patient records with an on-the-fly anonymization should limit these effects, but at the moment, it is still far from being usable in many institutions.

All these gaps finally lead to a usage gap. Clinicians rather use Google to search for images on the web than to search in patient records, where the access is limited by patient ID. User interface, speed and retrieval quality seem to have advantages with simple tools such as Google and this needs to be taken into account for new medical CBIR interfaces.

21.6.3 Future Developments

Image retrieval does have a bright future as does information retrieval in general. Information is produced in ever-increasing quantities and it also becomes increasing available, whether through patient record or via the Internet in teaching files or the scientific literature. One of the future challenges is to navigate in a meaningful way in databases of billions of images, allowing for effective and efficient retrieval, and at the same time a diversity in the results displayed and not simple duplicate images. Modern hospitals produced in the order of 100 GB or 120,000 images per day and few image retrieval systems could index this providing a high response speed.

By far the largest amounts of data produced in 3D and 4D tomographic data sets and there is still little research in this domain although a few approaches mainly for surface models do exist. To better integrate the entire amount of available information it also seems necessary to better integrate visual, textual and structured data retrieval into unique systems. Currently the research domains are totally separated, and a closer collaboration is necessary for working systems. The goal in the end should be to deliver the right information to the right people at the right time, and this information needs to include visual data.

Another important future task is the real evaluation of retrieval systems in clinical practice and thus in collaboration with clinicians to show their practical benefit. This is required to show that the impact of image retrieval can become real, and up to which level the impact can reach. Component-level evaluation is necessary to better understand what is currently working and what not and this not only on the system level but on the level of single components. Having all components accessible via standard interfaces could also

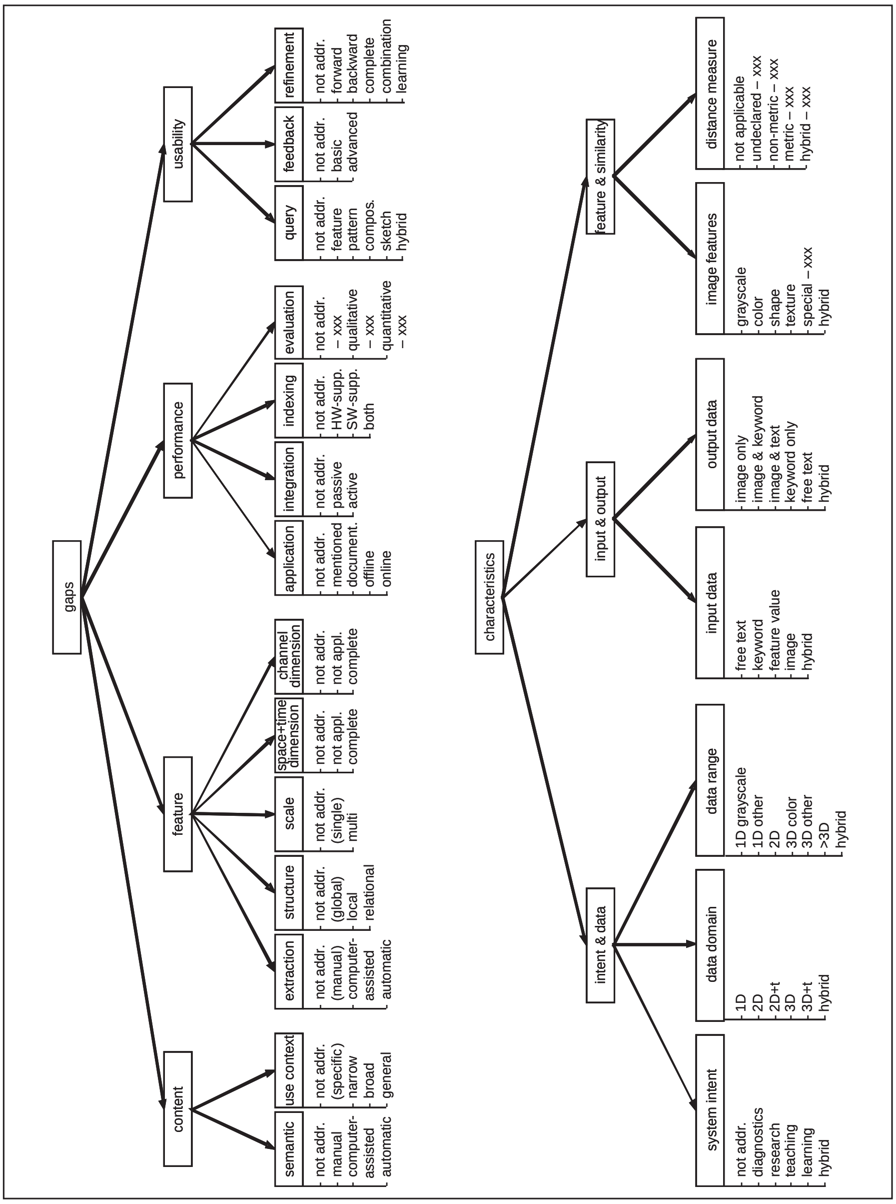

Fig. 21.13. Gaps in medical CBIR [80].

help to combine all possible components in a better way to really optimized the overall system performance.

References

1. Müller H, Michoux N, Bandon D, et al. A review of content-based image retrieval systems in medicine: clinical benefits and future directions. Int J Med Inform. 2004;73(1):1–23.
2. Tagare HD, Jaffe C, Duncan J. Medical image databases: a content-based retrieval approach. J Am Med Inform Assoc. 1997;4(3):184–198.
3. Chang SK, Kunii T. Pictorial data-base applications. IEEE Comput. 1981;14(11):13–21.
4. Enser PGB. Pictorial information retrieval. J Doc. 1995;51(2):126–170.
5. Gupta A, Jain R. Visual Information Retrieval. Commun ACM. 1997;40(5):70–79.
6. Eakins JP, Graham ME. Content-based image retrieval. Tech Report JTAP-039, JISC Technol Appl Program, Newcastle upon Tyne; 2000.
7. Venters CC, Cooper M. Content-based image retrieval. JTAP-054. J Inf Sci Res; 2000.
8. Smeulders AWM, Worring M, Santini S, et al. Content-based image retrieval at the end of the early years. IEEE Trans Pattern Anal Mach Intell. 2000;22(12):1349–1380.
9. Datta R, Joshi D, Li J, et al. Image retrieval: ideas, influences, and trends of the new age. ACM Compute Surv. 2008;40(2):1–60.
10. Chang NS, Fu KS. Query-by-pictorial-example. IEEE Trans Softw Eng. 1980;6(6):519–524.
11. Niblack W, Barber R, Equitz W, et al. Storage and retrieval for image and video databases: QBIC project: querying images by content, using color, texture, and shape. Proc SPIE. 1993;1908:173–187.
12. Pentland A, Picard RW, Sclaroff S. Photobook: tools for content-based manipulation of image databases. Int J Computer Vis. 1996;18(3):233–254.
13. Ma WY, Deng Y, Manjunath BS. Human vision and electronic imaging II: tools for texture- and color-based search of images. Proc SPIE. 1997;3016:496–507.
14. Carson C, Thomas M, Belongie S, et al. VISUAL: blobworld: a system for region-based image indexing and retrieval. Lect Notes Computer Sci. 1999;1614:509–516.
15. Belongie S, Carson C, Greenspan H, et al. Color- and texture-based image segmentation using EM and its application to content-based image retrieval. Proc Int Conf Comput Vis. 1998; p. 675–682.
16. Cox IJ, Miller ML, Omohundro SM, et al. Target testing and the PicHunter bayesian multimedia retrieval system. Adv Digit Libr. 1996; p. 66–75.
17. Squire DM, Müller H, Müller W, et al. Design and evaluation of a content-based image retrieval system. 2001; p. 125–151.
18. Sclaroff S, La Cascia M, Sethi S, et al. Unifying textual and visual cues for content-based image retrieval on the World Wide Web. Comput Vis Image Underst. 1999;75(1/2):86–98.
19. Rocchio JJ. Relevance feedback in information retrieval;. p. 313–323.

20. Müller H, Zhou X, Depeursinge A, et al. Medical visual information retrieval: state of the art and challenges ahead. Proc ICME. 2007; p. 683–686.
21. Geusebroek JM, van den Boogaard R, Smeulders AWM, et al. Color invariance. IEEE Trans Pattern Anal Mach Intell. 2001;23(12):1338–1350.
22. Milanese R, Cherbuliez M. A Rotation, translation and scale-invariant approach to content-based image retrieval. J Vis Commun Image Represent. 1999;10:186–196.
23. Weszka JS, Dyer CR, Rosenfeld A. A comparative study of texture measures for terrain classification. IEEE Trans Syst Man Cybern C Appl Rev. 1976;6(4):269–285.
24. Kuo WJ, Chang RF, Lee CC, et al. Retrieval technique for the diagnosis od solid breast tumors on sonogram. Ultrasound Med Biol. 2002;28(7):903–909.
25. Lu CS, Chung PC. Wold features for unsupervised texture segmentation. Proc Int Conf Pattern Recognit Lett. 1998; p. 1689–1693.
26. Squire DM, Müller W, Müller H, et al. Content-based query of image databases: inspirations from text retrieval. Pattern Recognit Lett. 2000;21(13-14):1193–1198.
27. Comaniciu D, Meer P, Foran D, et al. Bimodal system for interactive indexing and retrieval of pathology Images. Proc IEEE Workshop Appl Comput. 1998; p. 76–81.
28. Winter A, Nastar C. Differential feature distribution maps for image segmentation and region queries in image databases. Proc CBAIVL. 1999; p. 9–17.
29. Lucchese L, Mitra SK. Unsupervised segmentation of color images based on k-means clustering in the chromaticity plane. Proc CBAIVL. 1999; p. 74–78.
30. Loncaric S. A survey of shape analysis techniques. Pattern Recognit. 1998;31(8):983–1001.
31. Lapeer RJ, Tan AC, Aldridge R. A combined approach to 3D medical image segmentation using marker-based watersheds and active contours: the active watershed method. Lect Notes Computer Sci. 2002;2488:596–603.
32. Mikolajczyk K, Schmid C. A performance evaluation of local descriptors. IEEE Trans Pattern Anal Mach Intell. 2005;27(10):1615–1630.
33. Sivic J, Zisserman A. Video google: efficient visual search of videos. Toward Category-Level Object Recognition. 2006; p. 127–144.
34. Tversky A. Features of similarity. Psychol Rev. 1977;84(4):327–352.
35. Swain MJ, Ballard DH. Color indexing. Int J Computer Vis. 1991;7(1):11–32.
36. Aggarwal CC, Hinneburg A, Keim DA. On the surprising behavior of distance metrics in high dimensional space. Lect Notes Computer Sci. 2001;1973:420–434.
37. Hinneburg A, Aggarwal CC, Keim DA. What is the nearest neighbor in high-dimensional spaces? VLDB. 2000; p. 506–516.
38. Hanka R, Harte TP. Curse of dimensionality: classifying large multi-dimensional images with neural networks. CIMCSP. 1996;.
39. Vasconcelos N, Lippman A. A probabilistic architecture for content-based image retrieval. IEEE Comput Soc. 2000; p. 216–221.
40. Goh KS, Chang E, Cheng KT. Support vector machine pairwise classifiers with error reduction for image classification. J Assoc Comput Machinery. 2001; p. 32–37.
41. de Oliveira J, Machado A, Chavez G, et al. MammoSys: a content-based image retrieval system using breast density patterns. Comput Methods Programs Biomed. 2010; p. in press.

42. Westerveld T. Image retrieval: Content versus context. vol. 1. CID; 2000.
43. Salton G, Buckley C. Term weighting approaches in automatic text retrieval. Inf Process Manage. 1988;24(5):513–523.
44. Jain AK, Duin RPW, Mao J. Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell. 2000;22(1):4–37.
45. Lehmann T, Güld M, Thies C, et al. Content-based image retrieval in medical applications. Methods Inf Med. 2004;43(4):354–361.
46. Greenes RA, Brinkley JF. Imaging systems; 2000. p. 485–538.
47. Kulikowski C, Ammenwerth E, Bohne A, et al. Medical imaging informatics and medical informatics: opportunities and constraints. Methods Inf Med. 2002;41:183–189.
48. Sarvazyan AP, Lizzi FL, Wells PNT. A new philosophy of medical imaging. Med Hypotheses. 1991;36:327–335.
49. Rosset A, Ratib O, Geissbuhler A, et al. Integration of a multimedia teaching and reference database in a PACS environment. Radiographics. 2002;22(6):1567–1577.
50. Maloney K, Hamlet CT. The clinical display of radiologic information as an interactive multimedia report. J Digit Imaging. 1999;12(2):119–121.
51. Lowe HJ, Antipov I, Hersh W, et al. Towards knowledge-based retrieval of medical images. the role of semantic indexing, image content representation and knowledge-based retrieval. Proc Annu Symp Comput Appl Med Inform. 1998; p. 882–886.
52. Keysers D, Dahmen J, Ney H, et al. A statistical framework for model-based image retrieval in medical applications. J Electron Imaging. 2003;12(1):59–68.
53. Shyu CR, Brodley CE, Kak AC, et al. ASSERT: a physician-in-the-loop content-based retrieval system for HRCT image databases. Comput Vis Image Underst. 1999;75(1/2):111–132.
54. El-Naqa I, Yang Y, Galatsanos NP, et al. A similarity learnign approach to content–based image retrieval: application to digital mammography. IEEE Trans Med Imaging. 2004;23(10):1233–1244.
55. Wong JSJ, Zrimec T. Classification of lung disease pattern using seeded region growing. Proc AI. 2006; p. 233–242.
56. Depeursinge A, Iavindrasana J, Cohen G, et al. Lung tissue classification in HRCT data integrating the clinical context. In: Proc CBMS; 2008. p. 542–547.
57. Müller H, Rosset A, Vallée JP, et al. Integrating content-based visual access methods into a medical case database. Proc MIE. 2003;95:480–485.
58. Lehmann T, Goudarzi S, Linnenbrügger N, et al. Automatic localization and delineation of collimation fields in digital and film-based radiographs. Proc SPIE. 2002;4684:1215–1223.
59. Müller H, Kalpathy-Cramer J. Analyzing the content out of context - features and gaps in medical image retrieval. Int J Healthc Inf Syst Inform. 2009;4(1):88–98.
60. Névéol A, Deserno T, Darmonic S, et al. Natural language processing versus content-based image analysis for medical document retrieval. J Am Soc Inf Sci. 2009;60(1):123–134.
61. Welter P, Deserno T, Gülpers R, et al. Exemplary design of a DICOM Structured Report template for CBIR integration into radiological routine. Proc SPIE. 2010;7628:in press.
62. Lemke HU. PACS developments in europe. Comput Med Imaging Graph. 2002;27:111–120.

63. Qi H, Snyder WE. Content-based image retrieval in PACS. J Digit Imaging. 1999;12(2):81–83.
64. Bueno JM, Chino F, Traina AJM, et al. How to add content-based image retrieval capacity into a PACS. Proc CBMS. 2002; p. 321–326.
65. Welter P, Deserno T, Fischer B, et al. Integration of CBIR in radiological routine in accordance with IHE. Proc SPIE. 2009;7264:041–048.
66. Welter P, Deserno T, Gülpers R, et al. Exemplary design of a DICOM Structured Report template for CBIR integration into radiological routine. Proc SPIE. 2010;7628:in press.
67. Markkula M, Sormunen E. The challenge of image tetrieval, a workshop and symposium on image retrieval: searching for photos - journalists practices in pictorial IR. The British Computer Society; 1998.
68. Müller H, Despont-Gros C, Hersh W, et al. Health care professionals image use and search behaviour. 2006; p. 24–32.
69. Hersh W, Müller H, Gorman P, et al. Task analysis for evaluating image retrieval systems in the ImageCLEF Biomedical Image Retrieval Task. SOL. 2005;.
70. Lehmann T, Schubert H, Keysers D, et al. The IRMA code for unique classification of medical images. Proc SPIE. 2003;5033:440–451.
71. Depeursinge A, Vargas A, Platon A, et al. 3D case-based retrieval for interstitial lung diseases. Lect Notes Computer Sci. 2009; p. 18–25.
72. Fischer B WPGCGRDT Brosig A. Content-based image retrieval applied to bone age assessment. Proc SPIE. 2010;7624:in press.
73. Deserno T, Antani S, Long R. Content-based image retrieval for scientific literature access. Methods Inf Med. 2009;48(4):371–380.
74. Deserno T, Güld M, Plodowski B, et al. Extended query refinement for medical image retrieval. J Digit Imaging. 2008;21(3):280–289.
75. Hsu W, Antani S, Long L, et al. Int J Med Inform. 2009;78 Suppl 1), pages = S13-24, title = SPIRS: a Web-based image retrieval system for large biomedical databases.
76. Fischer B, Deserno T, Ott B, et al. Integration of a research CBIR system with RIS and PACS for radiological routine. Proc SPIE. 2008;6919:I41–I410.
77. Beretti S, Del Bimbo A, Pala P. Content–based retrieval of 3D cellular structures. IEEE Comput Soc. 2001; p. 1096–1099.
78. Kelly PM, Cannon M, Hush DR. Storage and retrieval for image and video databases III: query by image example: the CANDID approach. Proc SPIE. 1995;2420:238–248.
79. Aisen AM, Broderick LS, Winer-Muram H, et al. Automated storage and retrieval of thin-section CT images to assist diagnosis: system description and preliminary assessment. Radiol. 2003;228(1):265–270.
80. Deserno T, Antani S, Long R. Ontology of gaps in content-based image retrieval. vol. 22; 2008.
81. Iavindrasana J, Depeursinge A, Ruch P, et al. Design of a decentralized reusable research database architecture to support data acquisition in large research projects. Proc MedInfo. 2007;12.

List of Acronyms

CBIR	Content-Based Image Retrieval
CBVIR	Content-Based Visual Information Retrieval
CLEF	Cross Language Evaluation Forum
CT	Computed Tomography,
DICOM	Digital Imaging and Communications in Medicine , ,
ECG	Electrocardiography
EEG	Electroencephalography
GIFT	GNU Image Finding Tool,
GUI	Graphical User Interface
HARAG	Hierarchical Attributed Region Adjacency Graph
HL7	Health Level Seven
HSV	Hue-Saturation-Value
IHE	Integrating the Healthcare Enterprise
IRMA	Image Retrieval in Medical Applications, ,
LIDC	Lung Image Database Consortium
MedGIFT	Medical GIFT
MRI	Magnetic Resonance Imaging,
MRML	Multimedia Retrieval Markup Language
NCI	National Cancer Institute
NIH	National Institutes of Health,
NLM	National Library of Medicine
PACS	Picture Archiving and Communication System ,
PACS	Picture Archiving and Communication Systems
PET	Positron Emission Tomography
QBE	Query by Image Example
QBIC	Query by Image Content
RIS	Radiology Information System
ROI	Region of Interest
SPECT	Single Photon Emission Computed Tomography
SVM	Support Vector Machine

TREC	Text Retrieval Conference
UID	Unique Identifier

Index