Overview of the First Workshop on Medical Content–Based Retrieval for Clinical Decision Support at MICCAI 2009

Henning Müller[1,2], Jayashree Kalpathy–Cramer[3], Barbara Caputo[4], Tanveer Syeda-Mahmood[5], Fei Wang[5]

[1] Geneva University Hospitals and University of Geneva, Switzerland,
[2] University of Applied Sciences Western Switzerland (HES–SO), Sierre, Switzerland,
[3] Oregon Health and Sciences Univeristy (OHSU), Portland, OR, USA
[4] IDIAP research center, Martigny, Switzerland
[5] IBM Almaden research center, San Jose, CA, USA,

Abstract. In this paper, we provide an overview of the first workshop on Medical Content–Based Retrieval for Clinical Decision Support (MCBR–CDS), which was held in conjunction with the Medical Image Computing and Computer Assisted Intervention (MICCAI) conference in 2009 in London, UK. The goal of the workshop was to bring together researchers from diverse communities including medical image analyses, text and image retrieval, data mining, and machine learning to discuss new techniques for multimodal image retrieval and the use of images in clinical decision support. We discuss the motivation for this workshop, provide details about the organization and participation, discuss the current state–of–the–art in clinical image retrieval and the use of images for clinical decision support. We conclude with open issues and challenges that lie ahead for the domain of medical content–based retrieval.

1 Introduction

Diagnostic decision making (using images and other clinical data) is still very much an art for many physicians in their practices today due to a lack of quantitative tools and measurements [1]. Traditionally, decision making has involved using evidence provided by the patients data coupled with a physicians a priori experience of a limited number of similar cases [2]. With advances in electronic patient record systems and digital imaging, a large number of pre–diagnosed patient data sets are now becoming available [3]. These datasets are often multimodal consisting of images (X–ray, CT – Computed Tomography, MRI – Magnetic Resonance Imaging), videos and other time series, and textual data (free text reports and structured clinical data). Analyzing these multimodal sources for disease–specific information across patients can reveal important similarities between patients and hence their underlying diseases and potential treatments. Researchers are now beginning to use techniques of content–based retrieval to search for disease–specific information in imaging modalities to find supporting evidence for a disease or to automatically learn associations of symptoms and diseases [4] although already proposed over ten years ago [5, 6])

The diversity in medical imaging exams has risen enormously in the past 20 years as have the data amounts (The Geneva Hospital's Radiology department alone produced on average 114'000 images per day in 2009). Reading and interpreting multidimensional exams such as 4D data of a beating heart without computer support is extremely hard and requires much experience. At the same time all images are becoming available to clinicians via the electronic patient record [7]. This makes them available to potentially less experienced clinicians who relied on radiology reports beforehand and increases the risk of misinterpretations.

Benchmarking frameworks such as ImageCLEF[6] (Image retrieval track in the Cross–Language Evaluation Forum) have expanded over the past seven years to include large medical image collections for testing various algorithms for medical image retrieval [8–10]. This has made comparisons of several techniques for visual, textual, and mixed medical information retrieval as well as for visual classification of medical data possible based on the same data and tasks.

Image databases have also become available through several means such as the NCI[7] (National Cancer Institutes) and ADNI[8] (Alzheimer's Disease Neuroimaging Initiative). This lowers the entry burden to medical image analysis and should help to apply state–of–the–art techniques to medical imaging. Many open access journals such as BioMed Central[9] or Hindawi also make large amounts of the medical literature available that can then be indexed in tools such as ImageFinder[10] or MedSearch[11]. This search can include search for images by text and by visual means. Another tool that indexes openly accessible articles for the journals of the ARRS (American Roentgen Ray Society) is GoldMiner[12] [11]. All these tools and data can help to make the right information available to the right people at the right time to support healthcare applications including the use of visual data.

The first workshop on Medical Content–Based Retrieval for Clinical Decision Support (MCBR–CDS) was held at MICCAI (Medical Image Computing and Computer Assisted Interventions) 2009 in London, United Kingdom. The goal of the workshop was to bring together researchers from diverse communities including medical image analysis, text and image retrieval, data mining, and machine learning to discuss new techniques for multimodal image retrieval and the use of images in clinical decision support.

Content–based visual, textual, and multimodal information retrieval have been some of the promising techniques to help better manage the extremely

[6] http://www.imageclef.org/
[7] http://www.cancer.gov/
[8] http://www.adni-info.org/
[9] http://www.biomedcentral.com
[10] http://krauthammerlab.med.yale.edu/imagefinder/
[11] http://medgift.unige.ch:8080/MedSearch/faces/Search.jsp
[12] http://goldminer.arrs.org/

large amounts of visual information currently produced and used in most medical institutions around the world. As the topic has traditionally been close to rather applied research it has not yet been a primary topic at the MICCAI conference that rather concentrates on theoretically sound techniques in medical image processing. Still, to manage the increasingly large image archives in medical institutions, also the more theoretical researchers require to find the right images and use larger data sets, and on the other hand it is also time that image retrieval adopts some of the newer techniques of medical image processing, so bringing together the communities sounds like a logical step.

Submissions were proposed in the following principle areas of interests:

- data mining of multimodal medical data,
- machine learning of disease correlations from mining multimodal data,
- algorithms for indexing and retrieval of data from multimodal medical databases,
- disease model–building and clinical decision support systems based on multimodal analysis,
- practical applications of clinical decision support using multimodal data retrieval or analysis,
- algorithms for medical image retrieval or classification using the ImageCLEF collection.

A large variety of techniques were finally being submitted to the workshop. A selection of ten papers was taken that were presented orally at the workshop. Two high–quality invited speakers also presented their view on image analysis and retrieval for diagnosis support to round up the workshop.

2 Organization

MCBR–CDS was organized by an international set of researchers from the image retrieval, data mining and clinical decision support areas. The main organization was shared between Europe and the United States. The preface of these proceedings gives an overview of the organizers of the workshop and their roles in the process.

A sufficiently high number of reviewers were enlisted to help ensure the quality of presentations at the workshop. A total of over 30 experts from almost 20 countries helped in the review process.

3 Highlights of the Presentations

The workshop attracted researchers from industry and academia and from a multitude of domains, from computer science and imaging informatics to more clinically oriented groups. A total of 16 papers were submitted to the workshop. The double–blind review process included at least three external reviewers for each of the papers among the scientific community of over 30 international experts. All papers were then reread by the conference chairs and ranked based on the external reviews followed by a comparison of similarly scored papers. The ten best papers (60%) were chosen for oral presentation at the workshop using this process while the remaining papers were rejected. There was a good mix of papers between the image retrieval and clinical decision support domains as well as papers from a variety of clinical domains and using several imaging modalities. The discussions in the breaks and after the workshop underlined the interest of the presentations and the quality of the papers chosen to be presented at the workshop.

This section first presents an overview of the invited presentations and the goes into details into the three blocks of papers presented at the workshop on the topics image retrieval, clinical decision support and image annotation.

3.1 Invited Presentations

Two high quality key note presentations were given at the workshop to present external views on image retrieval and clinical decision support.

First, Dimitris Metaxas presented three projects of his research group in the context of image retrieval. The application domains from cardiac imaging to lung nodule detection, and combining sources for classification of pathology slides were presented. The importance of using high–quality data sets was highlighted to evaluate tools and algorithms on real, clinical, and thus naturally noisy data with all the difficulties that this implies. It was also shown that a variety of supporting techniques such as image segmentation in 3D and 4D are necessary for really detecting abnormal structures in the haystack of data that is often produced in medicine. Another important combination of sources that was presented was the classification of pathology slides using visual features and clinical data at the same time.

The second invited speaker was the clinician Scott Adelmann of Kaiser Permanente, the second largest hospital group in the United States. He first presented the challenges he sees in managing the currently 700 TB of image data stored by the group in connection with all the other clinical data . One of the biggest problems is the lack of structured data and thus a high–quality input for computerized tools to treat and interpret the data stored in past cases. He then presented several potential applications of image analysis and retrieval and their benefits. He clearly formulated his expectations to the image retrieval and multimodal analysis community in developing applications that are usable in practice and that can be evaluated in clinical settings.

3.2 Image Retrieval

The first set of papers describes applications of image retrieval in a variety of domains from microscopy, to CT images of the lung, MRI images of the brain and photographs of the skin, plus the use of visual and textual means combined for retrieval on the varied ImageCLEF database containing images from the scientific literature.

In [12], Andre et al. describe an application of endomicroscopic image retrieval using spatial and temporal features. They were motivated by the challenge of retrieving similar images using probe–based confocal laser endomicroscopy (pCLE), a recent imaging modality. Given a new image, they wanted to retrieve semantically similar images from a database of images annotated with expert diagnoses with the goal of aiding a physician trying to establish a diagnosis for the image. They extended the standard visual bag–of–words approach to content–based image retrieval (CBIR) by incorporating spatial and temporal information. Instead of using only salient points, they observed improved results by sampling more densely across the image. The incorporated spatial information using a co–occurrence of visual words. The temporal dimension was addressed by incorporating successive frames, a common approach in video retrieval. They also used image mosaicing to effectively increase the field of view (FOV). Using a leave–n–out cross validation technique, the authors demonstrated the efficacy of this approach by providing encouraging results on a small database of about a 1000 manually curated images. Particularly the visualization of the features of the salient points helped to understand this technique that is very often seen as a black box and thus hard to interpret.

Ballerini et al. [13] presented a query–by–example image retrieval for non–melanoma skin lesions. Dermatology, a visually oriented domain, has long been popular for computer–based image analysis and automated detection systems but relatively few CBIR systems have been described in the literature. The authors reiterate the need of CBIR systems in dermatology by highlighting the importance for being able to retrieve images that might be similar in appearance to cases on hand but having different diagnoses. This would be a useful tool for a clinician in identifying differential diagnoses. Their system focussed on five non-melanoma types of lesions including Actinic Keratosis, Basal Cell Carcinoma, Melanocytic Nevus, Squamous Cell Carcinoma and Seborrhoec Keratosis. They extracted a variety of color and texture features and used a genetic algorithm for the feature selection. Finally, they evualuted their system using precision recall curves. A particular emphasis was put on the database that contains fairly different lesions and different imaging techniques than most other computer–based tools to aid diagnosis in dermatology.

Depeursinge et al. [14] discussed their work in creating a computer–aided diagnosis (CAD) system that retrieves similar cases for interstitial lung diseases (ILDs) using 3–D high resolution computerized tomography images. The goal again is to assist clinicians, in this case emergency radiologists who often need to decide very quickly and are not experts in all application domains, in the process of establishing a diagnosis. They view automated tissue classication in the lung as being complementary to case–based retrieval from a computational as well as a user–centric standpoint. They began with a semi–automated segmentation of the lung, where only a single seed point is required. This was followed by texture based categorization of lung tissue using grey level histograms and wavelets as features, and a support vector machine classifier. Similar medical cases were then retrieved using a multimodal distance measure based on the volumes of segmented tissue groups as well as text–based clinical parameters extracted from the patient's health record including patient demographics, smoking status, laboratory results, and medical history. They compared the automated tissue segmentations with annotations performed by two expert radiologists. They achieved relatively good performance in tissue categorization and case–based retrieval by incorporating the clinical context of the patient. This is actually a critical aspect of image retrieval and decision support as the context of the patient (for instance age or smoking status) can affect the visual appearance of the imaged used for diagnosis. By combining the text–based clinical parameters with 3D imaging, the authors have created a helpful aid for diagnosis that is currently being tested in clinical practice.

In [15], Agarwal et al. describe a computerized image retrieval and diagnosis system for Alzheimer's disease (AD). Using the popular Alzheimer's Disease Neuroimaging Initiative (ADNI) MRI data–set, the authors have created a multilevel system for indexing and retrieving similar images based on textual or visual queries. They broach important trade–offs including retrieval versus classification, representation versus classification and representation versus retrieval in this paper. Their approach to retrieval is based on first classifying the query MRI image into one of three classes — AD, mild cognitive impairment and normal. This image classification is performed using Discrete Cosine Transform (DCT) features and an SVM classifier. Once the image has been classified, the ordered list of images returned to the user from within the class is identified. The authors provide results from an evaluation that demonstrates that the precision using the two step approach of reducing the search space first by classification and then retrieving similar images is better than retrieval using all possible images.

Rahman et al. [16] report on their results with multimodal image retrieval using the ImageCLEFmed 2008 database. They have approached the problem of retrieving relevant images from a collection of medical images and annotations using a multi–modal query expansion method that integrates both visual and textual features. Using a local feedback approach, they establish correlations between visual and text keywords. Their interactive system allows either keyword–based searches or query–by–example searches. The authors manually defined 30 local concept categories and using local color and textures features, used an SVM to train the image collection. The multi–modal similarity distance is then a weighted sum of visual and textual distances. The authors demonstrate an improved performance with the query expansion using the ImageCLEFmed 2008 database.

3.3 Clinical Decision Support

This set of papers dealt with the use of images and image retrieval techniques for clinical decision making.

The Breast Image Report and Data System (BI–RADS) [17] lexicon was developed by the American College of Radiology (ACR) to standardize the terminology in describing lesions in mammogram reports. It has been used extensively in breast cancer research for classifying mammograms and more recently, ultrasounds. In this paper, Zhang et al. [17] use an ensemble approach to classifying mammograms using the BI–RADS descriptors. Using an information–gain based approach to feature selection, they identified margin and shape to be most informative while noting that age and density could be left out without sacrificing performance. They first quantized the descriptors into coarser categories and then classified each category using the best classifier from an ensemble. The authors demonstrated that they achieved equivalent results to full fine–grained representations were obtained using the the coarse–grained descriptors of a subset of the BI–RADS features. They also noted that an ensemble learning approach outperformed the individual classifiers.

Quantitative gait analysis has been used in the diagnosis and treatment of a variety of illnesses in which the disease can have a profound impact on a patient's gait including Parkinson's disease, cerebral palsy and arthritis. The paper by Sen Köktas et al. [18] uses parameters to distinguish normal patients from those suffering from knee osteoarthritis (OA) using a set of 111 patients and 110 age–matched normal subjects. Using a commercial system, the researchers collected temporal changes of joint angles, joint moments, joint power, force ratios and time–distance parameters from four anatomical locations (pelvis, hip, knee and ankle) and in three motion planes (sagittal, coronal and traversal). For each of the 33 gait attributes, 51 samples were taken in the gait cycle, resulting in a 1653–dimensional feature vector. The number of features was first reduced using either a time–averaging or FFT technique. This was followed by a further reduction of dimensionality using the Mahalanobis distance, resulting in a final feature vector dimension of about 50. Finally, these features are used for classification using a set of linear and non–linear classifiers. The authors identified highly discriminative features using the Mahalanobis distance that corresponded well with those suggested by gait analysis experts and demonstrated good performance in the classification task using non–linear classifiers coupled with the use of time–averaging for dimensionality reduction. This paper could unfortunately not be presented at the workshop as the authors were unable to come at the very last moment.

Duchesne et al. [19] have described their approach to integrating information from various modalities including clinical, cognitive, genetic and imaging data to create a decision model to discriminate patients suffering from Alzheimer's disease from normal controls. The data are from the ADNI database and include the results of neuropsychological tests, quantitative hippocampal volumes obtained from imaging, and demographic and genetic risk factors including age, gender and APOE genotype. These data were integrated in the form of a binary string allowing the use of the Hamming distance for classification. The authors reported a 99.8% classification accuracy using 10–fold cross–validation.

3.4 Automatic Image Annotation

The final set of papers deals with the concept of automatic annotation of images based on visual appearance. Both papers in this set used the data from the Automatic Annotation task of ImageCLEF [8]. This annual medical image classification challenge consists of automatically annotating an image collection of more than 2000 X–rays given a training set of about 12'000 classified images. The images are classified using the IRMA (Image Retrieval in Medical Applications) [20] scheme with hierarchical labels for imaging modality, anatomical location, body system and image view. Some of the challenging aspects of this task include the highly unbalanced class memberships, significant intra–class visual dissimilarity as wells and inter–class similarity. The scoring system for the task was set up to reduce guessing when in doubt by penalizing an incorrect class more than an "unknown" class and penalizing mistakes lower down in the hierarchy less than errors closer to the top. The goals of this hierarchical weighting scheme was to force groups to add a confidence into the classification evaluation and give groups with good confidence score a better result. In 2009, the task was also made more difficult in that the distribution of the number of elements per class in training and testing data was deliberately not the same and even images from classes not occurring in the training data could be part of the test data.

Unay et al. [21] discuss their results of their participation in the ImageCLEFmed automatic annotation task. The main contribution of this paper was demonstrating that PCA–based local binary patterns used for the SVM classifier performed almost as well as the complete feature vector set, thereby enabling the use of a smaller dimensional feature vector and reducing computation time. This resulted in a 5–fold improvement in processing time and storage space requirements.

In [22], the authors propose a novel learning–based algorithm for medical image annotation that utilizes robust aggregation of learned local appearance evidences. This approach was applied to the task of automatically distinguishing the posteroanterior/anteroposterior (PA–AP) from the lateral (LAT) views of chest radiographs with the goal of integrating this as a post–processing module for computer–aided detection systems for both an in–house database as well as the ImageCLEF automatic annotation collection. The authors begin by demonstrating the within class variability found in appearance of radiographs of the chest and pelvis. The algorithm to identify the view of chest radiographs starts with the detection of landmarks using simultaneous feature selection and classification at different scales. Next, these landmarks are filtered using a sparse configuration algorithm to eliminate inconsistent findings. Finally a reasoning module identifies the final image class using these remaining landmarks. They demonstrate superior results in distinguishing PA–AP from LAT views for chest radiographs. An evaluation using a subset of the ImageCLEFmed classification collection was also performed using the most frequent classes of the database. The approach shows very good performance on a small number of classes with reasonably large difference between the classes. The scalability to a large number of classes was not attempted.

3.5 General Remarks on the Workshop

The workshop attracted researchers from a variety of clinical domains working on the challenges associated with image retrieval, primarily to aid with clinical decision support. Many different imaging modalities (x–ray, CT, MRI, endoscopy, photography, dermatology, microscopy) were presented as well as many anatomic regions (head, lung, colon, brain, skin, varied anatomic regions). Some traditional domains such as MRI image retrieval of the brain and CT image analysis of the lung were present as well as general image classification of X–rays. On the other hand several new domains were described such as the micro–endoscopy system described in [12]. Whereas most traditional approaches analyzed also 3D data mainly in 2D slices there were several approaches analyzing and calculating similarity directly in the 3D space [14, 15]. Time–series of images were tackled as well as combining visual features with clinical parameters

However, there are also several shortcoming that became apparent in this workshop. Few of the approaches are developed in close clinical collaboration meaning that a real use of the systems was basically not evaluated at all. Most often, publicly available databases or parts of them were used for evaluating specific aspects for the diagnosis aid process [15, 17, 19]. Sometimes, own databases were created with clinicians [13] but no full clinical evaluation was attempted. In general it is very hard to compare techniques and systems as often databases and setups are different for each system making comparison impossible. The next section will go deeper into the current challenges of image retrieval as diagnosis aid.

4 Open Issues and Challenges

The papers presented at the workshop show that the domain of medical image retrieval for clinical decision support is moving forward strongly. Still, some limitations could also be identified that require to be tackled in the near future to make image retrieval a tool usable for clinical practice. Notable, these challenges are:

- currently, only very few image retrieval systems have been developed in close collaboration with clinicians and have been evaluated clinically in a real workflow;
- purely visual systems do not take into account the clinical context and a combination seems necessary for a full decision support;
- standards for evaluation including data sets, ground truth ad criteria seem necessary to really be able to compare results on the same basis and show advances in the field.
- multidimensional data sets represent by far the largest amount of data produced in hospitals at the moment, i.e. 3D and 4D data as well as combinations of modalities, which have not been used for image retrieval, yet;
- the sheer size of PACS is not treated by any image retrieval system at the moment, particularly small Matlab prototypes will not scale to several million images.

These challenges will be detailed in the following sections.

4.1 Integration of Clinicians and Clinical Evaluation

One of the biggest challenges at the moment is bringing the theoretical applications that are often performed on small data sets and with MatLab prototypes towards real clinical applications. To our knowledge only a single evaluation has so far been performed with medical image retrieval [23] showing a clear improvement of diagnosis quality with the system use, particularly for little experienced radiologists. Most often, systems are developed in computer science departments far away from clinicians and with no direct collaboration other than an exchange of data. A close collaboration with clinicians and including frequent feedback from clinicians is necessary, which can become possible through direct collaborative projects.

Another important part in collaboration with clinicians is the analysis of the behavior of clinicians for example when using images or searching for them [24, 25]. Health care professionals have indicated that they would like to be able to restrict searches to a given modality, anatomy, or pathology of the image. However, the image annotations in on–line collections or teaching files do not always contain the information about the modality or anatomy. On the other hand, purely visual systems are not believed to be mature enough for image retrieval for images with specific pathological findings, especially for image collections containing a variety of image modalities and pathologies. One thing mentioned was also the search for visually similar images but with different diagnosis, to illustrate teaching and also to differential diagnoses. One big problem with visual retrieval currently is that the formulation of information needs with visual means is far from easy.

Another important aspect for medical image retrieval is the notion of relevance in medical image search. This is somewhat researched for still images but for 3D or 4D data sets this is not clear. Past tests have also shown that this is person–dependent, task dependent, and depends strongly on the knowledge of the searching person on a particular domain [9].

Another way to convince clinicians is to have a clear proof of retrieval quality as few people would want to work with systems that can not show a certain quality level. To show such performance, standard data sets are extremely important [26] and also a methodology to evaluate several systems based on the same means.

4.2 Multimodal Data Treatment and Information Fusion

Purely visual techniques may not be sufficient for most clinical applications. In medicine, visual information taken alone is less meaningful than the same images viewed in the context of the patient and the clinical environment. We believe that pure CBIR methods in medicine have not lived up to expectations due to their inability to incorporate context. No medical doctor would diagnose based on images, only, as the context carries much of the necessary information to interpret

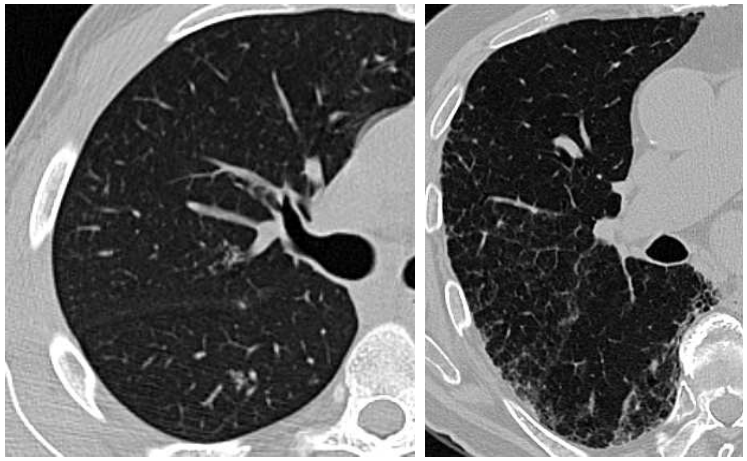

Fig. 1. Comparing the healthy tissue of a 25–year-old and an 88-year–old person shows the important differences in grey level and texture.

images. Image retrieval in medicine needs to evolve from purely visual retrieval to a more holistic, case–based approach that incorporates various multimedia data sources. These include multiple images, free text, structured data, as well as external knowledge sources and ontologies.

The semantic gap poses one of the major challenges in creating a useful image retrieval engine. Smeulders [27] identified the *semantic gap* as the lack of coincidence between the information that one can automatically extract from the visual data and the interpretation that the same data have for a given user in a given situation. In medical images, the semantic gap can manifest itself as a difference between the image and the interpretation of the image by the medical doctor including anamnesis, lab results, and potentially other exams. The same image may be interpreted differently depending on the medical doctor, his training, expertise, experience, and the context of the image acquisition and the patient. Such coincidences between content and contextual data have already been described in the non–medical field in [28] as well.

Effective clinical image retrieval systems can be used as a diagnostic aid. By allowing clinicians to view similar images contextually, they receive assistance in the diagnostic decision–making process by accessing knowledge of older cases. When being pro–active in this process missing data such as lacks in the anamnesis can be pointed out by the system and the clinician can directly ask the questions with the highest clinical information gain to the patient or order the corresponding lab examinations, as proposed by a computerized decision aid. This of course requires much more knowledge about a particular domain and the interrelations of the clinical data.

Examples for the importance of combining the textual and visual data are manifold. Figure 1 shows as an example the healthy lung of a 25–year–old and an

88-year–old. The lung of the 88–year–old shows several pre–fibrotic lesions and has a slightly altered grey value. Inverting age on the image of the 88–year–old would mean that the persons is not healthy but has a severe problem. Another example is the importance of the goal of the imaging study as it provides the context in which the image is to be viewed. CT images have a high dynamic range. The window/level settings must be set appropriately to provide detail and contrast for the organ of interest in the imaging study. Often, images are stored in JPEG for teaching and conference presentations and also in this case the right level/window setting when transferring the image is crucial. Whereas CT images usually have 1000–4000 grey levels, jpeg images only have 256, and most computer screens to not manage to show more than 256 different grey levels, either. Looking at chest CTs of the mediastinum or of the lungs would require totally different level window settings than looking at the lung, although the exactly same regions is show on the image.

Yet another example deals with the ability to incorporate patient history into the context in which the images are evaluated. In patients with lung cancer, radiation therapy is often delivered to the chest as part of the treatment plan. Many of these patients develop lung inflammation, known as pneumonitis. Some patients also develop radiation fibrosis, a scarring of the lungs. This can be mistaken for other interstitial lung diseases if the context of the patient is ignored in viewing subsequent scans of the chest. There are numerous other examples where the role of context is vital in the use of imaging studies for diagnosis and treatment. The lesions of multiple sclerosis (MS) can mimic a brain tumor and vice versa. A radiologist who is not aware of the clinical history of the patient as having MS can misdiagnose a suspicious lesion on an MRI.

All these examples underline that images can basically not be viewed correctly without clinical information and albeit this, most of the medical image retrieval systems currently ignore clinical data other than images totally.

4.3 Treating Extremely Large Databases

In the same way as for general Internet search engines one of the most important aspects for medical image retrieval systems is to be as complete as possible and as large as possible. Medical image repositories have multiplied over recent years with tools such as MyPACS[13] and MDPixx[14]. Even a standard to interconnect digital teaching files exists with MIRC[15] (Medical Imaging Resource Center). With the scientific literature another large body of knowledge that includes many medical images has become accessible [29] and is increasingly integrated with visual retrieval systems. Several web interfaces such as Goldminer[16] allow access to many hundred thousand images and these numbers are very likely to increase strongly and quickly making available for information search ever larger amounts of medical knowledge including images.

13 `http://www.mypacs.net/`
14 `http://www.mdpixx.org/`
15 `http://mirc.rsna.org/`
16 `http://goldminer.arrs.org/`

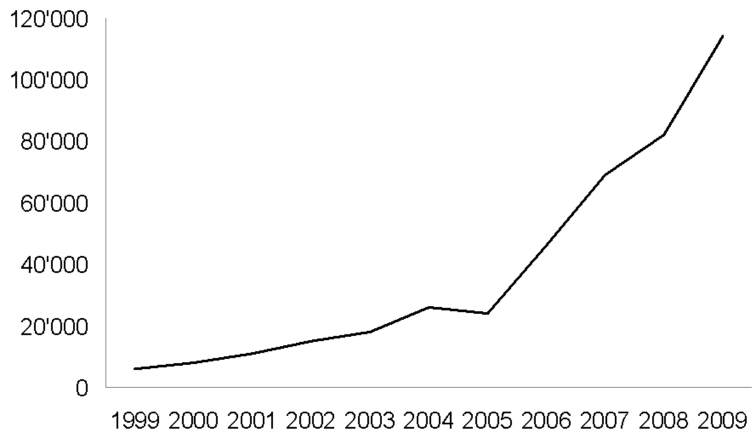

Fig. 2. The daily image production in the Radiology Department of the Geneva University hospitals has increased enormously over the past ten years.

The daily image production at the Geneva University Hospitals' radiology department (see Figure 2) also shows that internal data sets have grown exponentially and are continuing to grow at these rates. Multi–slice scanners and combinations of modalities such as PET/CT are some of the largest data producers. Large University hospitals often produce in access of 100 GB per day. Indexing the entire PACS for image retrieval in clinical routine has been proposed many times [30, 31] but to our knowledge not a single implementation has been performed up to now. Blocking parts include the legal aspects of accessing patient data but also the sheer amount of information that will require new index structures to cope with the several million images thus potentially available. Large benchmarking databases currently contain rarely more than 100'000 images, which is often already on the limit for prototypes in MatLab or systems that contain the features in main memory. Indexing and thus making accessible extremely large data sets still contains many challenges, also in reaching interactive response times.

4.4 Treating Multidimensional Data Sets

Currently, image retrieval is most often limited to single two–dimensional images. CT, MRI and also combined modalities such as PET/CT and PET/MRI are by far the largest production of data in hospitals (including time series of such data, so 4D data sets that have never been used for image retrieval, so far). Only very few systems analyze these multidimensional data sets directly.

Already viewing these data sets creates the challenging parts as humans are good in qualitative analysis but can usually not remember more than a few things at a time (from 3–7 according to the psychological literature [32, 33]). This means that viewing such multidimensional data with many aspects requires much

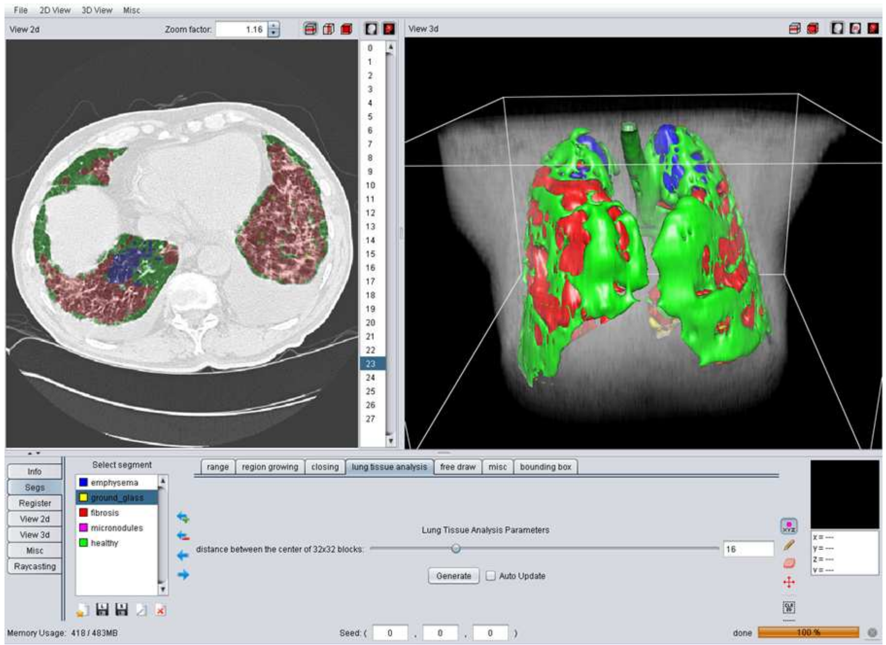

Fig. 3. An interface visualizing classified tissue in a 2D and a 3D view.

experiences and puts a stress on the clinicians by having to integrate very large amounts of data. Analyzing the data with diagnosis aid tools and highlighting potentially abnormal regions can at least reduce the stress through at least a somewhat second opinion.

Figure 3 shows a simple image retrieval system that visualizes the obtained cases directly in 3D and shows them to the clinicians. This is interface is a web interface (thus easy to integrate into clinical applications) using Java3D and particularly the YaDiV[17] (Yet another DICOM Viewer) system for visualization. The interface allows the clinicians to navigate directly in the 3D data if necessary but to concentrate on areas with abnormal parts in more details by automatically pre–classifying the entire lung tissue ahead of viewing.

5 Conclusions

This paper shows the advances that image retrieval has made over the past years through the ten presentations given at the workshop on Medical Content–Based Retrieval for Clinical Decision support. It can be seen that image retrieval is leaving the paradigm of taking image similarity from single images and that an integration of data from 3D data sets, clinical data, and also temporal data of time series has started. Standard data sets are available in some domains and

[17] http://www.welfenlab.de/en/research/fields_of_research/yadiv/

are used increasingly, albeit not always in the same ways. Data sets need to contain ground truth and clear criteria of success to allow for a real comparison of techniques to measure progress.

Systems also need to get closer to clinicians and show their potential with, at least, small trials. To do so, the integration of clinical with the visual data seem necessary. Very large amounts of clinical data are available and their integration into clinical applications using techniques from image retrieval seems necessary. Such large amounts are often accessible but they need to be integrated and accessible quickly.

Image retrieval in medicine needs to evolve from purely visual image retrieval to a more holistic, case–based approach that incorporates various multimedia data sources and thus the context in which the images were taken. This needs to include multiple images, free text, structured data as well as external knowledge sources and ontologies. All these data can consequently be integrated with literature databases such as Goldminer to give a clinician access to the right information (peer–reviewed literature, past cases with treatment and outcomes) at the right time and in the right format.

6 Acknowledgements

This work was partially funded by the Swiss National Science Foundation (FNS) under contract 205321–109304/1, the American National Science Foundation (NSF) with grant ITR–0325160, the TrebleCLEF project and Google. We would like to thank the RSNA for supplying the images of their journals Radiology and Radiographics for the ImageCLEF campaign. Thanks also to the HES–SO for funding the BeMeVIS project. A big Thank You also to the two invited speakers who helped to make the workshop a success. Thanks belong equally to IBM for supporting the workshop organization. Particular thanks also go to the MICCAI workshop organizer Daniel Elson for well organizing all logistical and practical aspect of the workshop and for remaining calm when we not respecting deadlines or forgot to respond to important mails.

References

1. Kohn, L.T., Corrigan, J.M., Donaldsen, M.S.: To Err is Human – Building a Safer Health System. National Aacademic Press, Washington DC, USA (1999)
2. Aamodt, A., Plaza, E.: Case–based reasoning: Foundational issues, methodological variations, and systems approaches. Artificial Intelligence Communications **7**(1) (1994) 39–59
3. Safran, C., Bloomrosen, M., Hammond, W.E., Labkoff, S., Markel-Fox, S., Tang, P.C., Detmer, D.E.: Toward a national framework for the secondary use of health data: An american medical informatics association white paper. Methods of Information in Medicine **14** (2007) 1–9
4. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content–based image retrieval systems in medicine – clinical benefits and future directions. International Journal of Medical Informatics **73**(1) (February 2004) 1–23

5. Tagare, H.D., Jaffe, C., Duncan, J.: Medical image databases: A content–based retrieval approach. Journal of the American Medical Informatics Association **4**(3) (1997) 184–198
6. Lowe, H.J., Antipov, I., Hersh, W., Arnott Smith, C.: Towards knowledge–based retrieval of medical images. The role of semantic indexing, image content representation and knowledge–based retrieval. In: Proceedings of the Annual Symposium of the American Society for Medical Informatics (AMIA), Nashville, TN, USA (October 1998) 882–886
7. Haux, R.: Hospital information systems — past, present, future. International Journal of Medical Informatics **75** (2005) 268–281
8. Müller, H., Kalpathy-Cramer, J., Kahn Jr., C.E., Hatt, W., Bedrick, S., Hersh, W.: Overview of the ImageCLEFmed 2008 medical image retrieval task. In Peters, C., Giampiccolo, D., Ferro, N., Petras, V., Gonzalo, J., Peñas, A., Deselaers, T., Mandl, T., Jones, G., Kurimo, M., eds.: Evaluating Systems for Multilingual and Multimodal Information Access – 9th Workshop of the Cross-Language Evaluation Forum. Volume 5706 of Lecture Notes in Computer Science., Aarhus, Denmark (September 2009) 500–510
9. Müller, H., Kalpathy-Cramer, J., Eggers, I., Bedrick, S., Said, R., Bakke, B., Kahn Jr., C.E., Hersh, W.: Overview of the 2009 medical image retrieval task. In: Working Notes of CLEF 2009 (Cross Language Evaluation Forum), Corfu, Greece (September 2009)
10. Müller, H., Deselaers, T., Lehmann, T., Clough, P., Kim, E., Hersh, W.: Overview of the ImageCLEFmed 2006 medical retrieval and annotation tasks. In: Working Notes of the 2006 CLEF Workshop, Alicante, Spain (Septermber 2006)
11. Kahn Jr., C.E., Thao, C.: Goldminer: A radiology image search engine. American Journal of Roentgenology **188** (2008) 1475–1478
12. Andre, B., Vercauteren, Perchant, A., Buchner, A., Wallace, M., Ayache, N.: ntroducing space and time in local feature-based endomicroscopic image retrieval. In Caputo, B., Müller, H., Syeda Mahmood, T., Kalpathy-Cramer, J., Wang, F., Duncan, J., eds.: MCBR–CDS 2009: Medical Content–based Retrieval for Clinical Decision Support. Volume 5853 of Lecture Notes in Computer Science., Springer (September 2009)
13. Ballerini, L., Fisher, R., Rees, J.: A query–by–example content–based image retrieval system of non–melanoma skin lesions. In Caputo, B., Müller, H., Syeda Mahmood, T., Kalpathy-Cramer, J., Wang, F., Duncan, J., eds.: MCBR–CDS 2009: Medical Content–based Retrieval for Clinical Decision Support. Volume 5853 of Lecture Notes in Computer Science., Springer (September 2009)
14. Depeursinge, A., Vargas, A., Platon, A., Geissbuhler, A., Poletti, P.A., Müller, H.: 3D case–based retrieval for interstitial lung diseases. In Caputo, B., Müller, H., Syeda Mahmood, T., Kalpathy-Cramer, J., Wang, F., Duncan, J., eds.: MCBR–CDS 2009: Medical Content–based Retrieval for Clinical Decision Support. Volume 5853 of Springer Lecture Notes in Computer Science., Springer Lecture Notes in Computer Science (September 2009)
15. Agarwal, M., Mostafa, J.: Image retrieval for alzheimer's disease detection. In Caputo, B., Müller, H., Syeda Mahmood, T., Kalpathy-Cramer, J., Wang, F., Duncan, J., eds.: MCBR–CDS 2009: Medical Content–based Retrieval for Clinical Decision Support. Volume 5853 of Lecture Notes in Computer Science., Springer (September 2009)
16. Rahman, M., Antani, S.: Multi–modal query expansion based on local analysis for medical image retrieval. In Caputo, B., Müller, H., Syeda Mahmood, T., Kalpathy-Cramer, J., Wang, F., Duncan, J., eds.: MCBR–CDS 2009: Medical Content–based

Retrieval for Clinical Decision Support. Volume 5853 of Lecture Notes in Computer Science., Springer (September 2009)
17. Zhang, Y., Tomuro, N., Furst, J., Raicu, D.S.: Using bi–rads descriptors and ensemble learning for classifying masses in mammograms. In Caputo, B., Müller, H., Syeda Mahmood, T., Kalpathy-Cramer, J., Wang, F., Duncan, J., eds.: MCBR–CDS 2009: Medical Content–based Retrieval for Clinical Decision Support. Volume 5853 of Lecture Notes in Computer Science., Springer (September 2009)
18. Sen Köktas, N., Duin, R.P.W.: Statistical analysis of gait data to assist clinical decision making. In Caputo, B., Müller, H., Syeda Mahmood, T., Kalpathy-Cramer, J., Wang, F., Duncan, J., eds.: MCBR–CDS 2009: Medical Content–based Retrieval for Clinical Decision Support. Volume 5853 of Lecture Notes in Computer Science., Springer (September 2009)
19. Duchesne, S., Crépeaut, B., Frisoni, G.: Knowledge–based discrimination in alzheimer's disease. In Caputo, B., Müller, H., Syeda Mahmood, T., Kalpathy-Cramer, J., Wang, F., Duncan, J., eds.: MCBR–CDS 2009: Medical Content–based Retrieval for Clinical Decision Support. Volume 5853 of Lecture Notes in Computer Science., Springer (September 2009)
20. Lehmann, T.M., Schubert, H., Keysers, D., Kohnen, M., Wein, B.B.: The IRMA code for unique classification of medical images. In Huang, H.K., Ratib, O.M., eds.: Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation. Volume 5033 of Proceedings of SPIE., San Diego, California, USA (May 2003) 440–451
21. Unay, D., Soldea, O., Ekin, A., Cetin, M., Ercill, A.: Automatic annotation of x–ray images: A study on attribute selection. In Caputo, B., Müller, H., Syeda Mahmood, T., Kalpathy-Cramer, J., Wang, F., Duncan, J., eds.: MCBR–CDS 2009: Medical Content–based Retrieval for Clinical Decision Support. Volume 5853 of Lecture Notes in Computer Science., Springer (September 2009)
22. Tao, Y., Peng, Z., Jian, B., Xuan, J., Krishnan, A., Zhou, X.S.: Robust learning based annotation of medical radiographs. In Caputo, B., Müller, H., Syeda Mahmood, T., Kalpathy-Cramer, J., Wang, F., Duncan, J., eds.: MCBR–CDS 2009: Medical Content–based Retrieval for Clinical Decision Support. Volume 5853 of Lecture Notes in Computer Science., Springer (September 2009)
23. Aisen, A.M., Broderick, L.S., Winer-Muram, H., Brodley, C.E., Kak, A.C., Pavlopoulou, C., Dy, J., Shyu, C.R., Marchiori, A.: Automated storage and retrieval of thin–section CT images to assist diagnosis: System description and preliminary assessment. Radiology **228**(1) (July 2003) 265–270
24. Hersh, W., Jensen, J., Müller, H., Gorman, P., Ruch, P.: A qualitative task analysis for developing an image retrieval test collection. In: ImageCLEF/MUSCLE workshop on image retrieval evaluation, Vienna, Austria (2005) 11–16
25. Müller, H., Despont-Gros, C., Hersh, W., Jensen, J., Lovis, C., Geissbuhler, A.: Health care professionals' image use and search behaviour. In: Proceedings of the Medical Informatics Europe Conference (MIE 2006). IOS Press, Studies in Health Technology and Informatics, Maastricht, The Netherlands (August 2006) 24–32
26. Vannier, M.W., Summers, R.M.: Sharing images. Radiology **228** (2003) 23–25
27. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content–based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and machine Intelligence **22**(12) (December 2000) 1349–1380
28. Westerveld, T.: Image retrieval: Content versus context. In: Recherche d'Informations Assistée par Ordinateur (RIAO'2000) Computer–Assisted Information Retrieval. Volume 1., Paris, France, CID (April 2000) 276–284

29. Müller, H., Kalpathy-Cramer, J., Kahn Jr., C.E., Hersh, W.: Comparing the quality of accessing the medical literature using content–based visual and textual information retrieval. In: SPIE Medical Imaging. Volume 7264., Orlando, Florida, USA (February 2009) 1–11
30. Bueno, J.M., Chino, F., Traina, A.J.M., Traina, C.J., Azevedo-Marques, P.M.: How to add content–based image retrieval capacity into a PACS. In: Proceedings of the IEEE Symposium on Computer–Based Medical Systems (CBMS 2002), Maribor, Slovenia (2002) 321–326
31. Qi, H., Snyder, W.E.: Content–based image retrieval in PACS. Journal of Digital Imaging **12**(2) (1999) 81–83
32. Cowan, N.: The magical number 4 in short–term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences **24**(1) (2001)
33. Miller, G.A.: The magical number seven plus or minus two: Some limits on our capacity for processing information. The Psychological Review **63** (1956) 81–97