Putting the Content Into Context: Features and Gaps in Image Retrieval

Henning Müller, University of Geneva and University of Applied Sciences, Switzerland
Jayashree Kalpathy-Cramer, Oregon Health and Science University, USA

ABSTRACT

150 words or less.

Keywords: Please insert keywords

INTRODUCTION

Digital multimodal archives have become ubiquitous with the rapid growth of the Internet, available computing power, and other technological advances, leading to immense amounts of digital multimodal data generation in the information society. Most common forms of such data include structured data, free text, audio, images, and videos, and of course combinations of all these. The need for semi- or fully-automatic means of organizing massive databases containing structured and unstructured components in this multimodal environment has exploded with the generation speed of such databases greatly exceeding the anticipated rates.

Images in Clinical Practice

Medical images have become a significant component of clinical practice and research (1). Due to advances in medical imaging technology, vast quantities of medical images covering a large variety of conditions are produced and stored. This variety is steadily growing with new imaging technologies developing (new contrast agents, higher resolutions, and thinner slices) and combinations of modalities such as PET/CTs making it even harder for clinicians
to really understand all available information sources. Combining all the available information sources for a single patient is even harder as psychological literature shows clearly that humans can only integrate a fairly small number of information sources, from 3-7 depending on the tests (2, 3). The accessibility of these data in the electronic patient record for all clinicians makes the situation even worse as not only specialists access the data but all clinicians (4). Undoubtedly, the effective management of such visual data, including x-ray images, computed tomography (CT) scans, magnetic resonance imaging (MRI), and non-radiology imaging sources, is imperative to maximize the utility of the collected images and to maximize the accuracy and efficiency of the health services. Images convey more information to the medical researcher or practitioner than can be abstracted in a brief report or annotation. Critical diagnostic and interventional decisions are based on the digital images acquired from a particular patient and often assessed in comparison with historical cases that are individually or institutionally accumulated such as in the Casimage system (5).

An effective medical image retrieval system can not only play a crucial role for clinical care, but it can also contribute greatly to medical research by allowing scientists to identify images of relevant cases more accurately and efficiently. It can prove to be extremely beneficial for medical students, as well as for patients and the general public to identify information relevant to their health related search. However, only a few studies (6) have looked at the user-behavior of image retrieval system users. This study noted that many clinicians store reference images from past cases, often on their personal computers, and also that most often images are searched for by pathology and not anatomic region or modality that are often implemented for image classification.

**Image Retrieval Techniques**

Traditionally, image retrieval systems have been text-based (7), relying on the annotations or captions associated with the images as the input to the retrieval system. This technique has many limitations as 1) the annotations are often subjective and context sensitive; 2) the task of manual indexing is labor and time intensive and also error prone; 3) there is far more information in an image than can be abstracted using a limited number of words.

In clinical applications, most medical personnel retrieve images using a patient or study identifier in the Picture Archival and Communication Systems (PACS). Thus, most image accesses in this scenario are purely patient-centered and the important knowledge that is stored in cases of other patients is not at all taken into account. However, the need for retrieval systems that offer features beyond the capabilities of standard PACS systems has been expressed many times (8,9,10). These include searching by anatomic region, pathology, visual similarity, and multi-modality combined to find similar cases and case-based searching capability. Recent results suggest that a multimodal approach combining visual and textual features is promising and usually leads to best overall results (11,36).

Visual retrieval results can be used to re-rank images retrieved through text and this can significantly improve early precision (38), in the example the mixed run had a P5 of 0.55 whereas the best textual system based on MAP had a P5 of 0.45; when sorting by MAP it is the other way around with the first system obtaining a MAP of 0.15 and the second of 0.21. Another approach is described in (37), where clinical attributes are included into the classification of regions of interest in lung CT images. This showed to improve classification results from 84% to 91%. Most clinical features were complementary to visual features but a few strong correlations were also found. Most other approaches currently use linear combinations of visual and textual retrieval and then combine the results. Usually, much care needs to be taken with respect to how to combine results as the. Not all combined systems have better results than text retrieval alone. More on this subject can also be read in (39).
Content-Based Image Retrieval

Advances in computer vision have led to methods for using the image itself as the search entity since the early 1980s (12). The query-by-example paradigm can be used in cases where the user cannot express his/her information need appropriately in a semantic fashion or where the system does not allow searching for these semantic expressions (for example: “Show me lung x-rays that look similar to tuberculosis”). This can arise if the searcher is not familiar with the findings in a given image as in the case of a clinician with an uncertain diagnosis, or a German speaking researcher searching for images in an English collection, or if the concept of the image cannot be abstracted easily.

Content-based image retrieval (CBIR) emerged as a natural consequence of this need and has evolved significantly in the past decade. In content-based image retrieval, the visual information from the image is mathematically abstracted and compared to similar abstractions of all images in the database. Ordered lists of images that are visually most similar to the sample image are presented to the user. Given a similarity metric, a query image is compared to each element of the database to identify a sorted list of the most visually similar elements that is returned to the user with the expectation that the features and the metric used match the visual expectations of the user.

Features used for CBIR can be local (i.e. concerning only a small region of the image) or global (rather about the general layout of an image). They most often include descriptions based on the color, shape, and texture of the images. These can include color features such as histograms, texture features including those based on wavelets, co-occurrence matrices, shape features, salient points, patch histograms, and many others.

Evaluation in Image Retrieval

To be able to compare current techniques based on the same datasets and tasks, several initiatives have started in the past few years. Previously, the identification of good or promising techniques was almost impossible as everyone used different datasets and evaluation methodologies (13). Several examples for evaluation based on unrealistic datasets or tasks can be found (14,15). The first active initiative was most likely the Benchathlon, identifying important evaluation constraints and common data sets. The most successful is surely TRECVID (16) with over 100 subscribing research groups in 2008. ImageCLEF, has started as part of CLEF (Cross Language Evaluation) in 2004, and since 2005 a medical image retrieval benchmark was added (17). Other image retrieval benchmarks include ImageEVAL and INEX MM (18).

Challenges in Current Medical Image Retrieval

General purpose image retrieval in most commercial applications such as Google or Yahoo! images is still accomplished by means of the textual annotation associated with the image, and only very specific techniques such as the detection of faces in images are currently included in these search engines. This is also true for the on-line medical image search engines such as Goldminer or while searching on-line databases of cases such as MyPACS or MIRC (Medical Imaging Resource Center). However, these systems are limited by the quality (and sometimes also quantity) of the annotations. The ability to search for visually similar images can be valuable in several scenarios, for example when a new case is available but no clear idea of the diagnosis exists. For education, the search for visually similar images with varying diagnosis is also important and can currently not be covered with any textual means.

CBIR systems in medicine are starting to make inroads, although on a limited and primarily research basis (19). However, most existing medical image retrieval techniques significantly lag their textual counterparts in their ability to capture the semantic essence of the user’s query (8). Abstracting the semantic essence of an image remains a challenging research
topic. The utility of purely visual CBIR systems can be limited in clinical practice due to the semantic and sensory gaps (20); several other challenges for image retrieval are also defined and classified in (21). In this paper, we mainly describe the content gap that actually includes the clinical context but also the usability and feature gaps are part of the problems described in this paper.

**Sensory Gap**

The early years of CBIR have been reviewed in a relatively comprehensive fashion by Smeulders et al. (20). The sensory gap was identified as the difference between “the object in the world and the information in a computational description derived from a recording of the scene”. A manifestation of the sensory gap in medical images is in the differences between the actual tumor in the physical world and how it is imaged under various modalities (e.g., CT or MRI) and views (prone or supine). X-rays as 2D representations of a 3D world with many overlapping structures have an extremely high loss concerning the sensory gap.

**Semantic Gap**

The semantic gap poses one of the largest challenges in creating a useful image retrieval engine. Smeulders et al. (20) 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 amnesia, 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.

Research on trying to close the semantic gap is an ongoing quest (22,23) in general image retrieval. Automatically extracted low level visual features do not necessarily correspond to high level concepts that a user has in his mind for searching. In CBIR, the semantic gap between low-level image features and high-level concepts that an image represents to a given user remains a challenge as does the issue of scalability of solutions to various sources of variability in broad-context image databases. The probability distribution of high level concepts given the low-level features of an image, or multimodal data, in general, is highly dependent on the purpose of the user.

**Other Challenges and Deficiencies in Image Retrieval**

Image retrieval, other than by patient or series ID, has not gained much traction in clinical practice. Clinical image retrieval systems need to be adapted to meet to domain and user-specific requirements and be integrated within the workflow to provide maximum benefit to clinicians.

In comparing image retrieval to text retrieval, Smeulders et al. note the lack of a sensory gap in text retrieval (20) and point out the difference between the semantic gap in text retrieval (between keywords to full text) to that in image retrieval. The differences in semantic and sensory gaps between textual and visual retrieval may shed some light on why image retrieval systems currently do not perform as well as their textual counterparts.

Müller et al. have performed an extensive review of the use of image retrieval in medicine (8). Image retrieval in medicine is most commonly performed within the area of PACS systems, where the images are retrieved using either the patient or study ID. However, Müller et al. advocate the introduction of content-based methods and assert that these can provide a useful functionality to existing systems very complementary to text-based information retrieval. Teaching, research, and diagnostics are identified as the three primary domains for applying image retrieval. An important differential analysis application that purely visual (or content-based) image retrieval will contribute
to is identified as follows: “visual features do not only allow the retrieval of cases with patients having similar diagnoses but also cases with visual similarity but different diagnoses.” This can be a very useful scenario in teaching, for example.

However, most of the clinicians interviewed in (24) do not believe that the CBIR systems in medicine are ready to be used in a clinical setting. They identified “recommendations for search techniques that do not exist but are regarded as very useful: search by pathology; search by anatomic region; search by visual similarity; search by multimodality combined to find similar cases; indexation of the entire PACS by keywords regarding the pathology.”

Users (24) 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. The ImageCLEFmed experience has clearly demonstrated that combining visual and textual methods can offer benefit (17,38). Fusion of multimodal retrieval techniques is a research topic that is of increasing importance (25), not only in the medical domain (26).

**THE ROLE OF CONTEXT IN MEDICAL IMAGE RETRIEVAL**

Computer vision generally concentrates on purely visual problems. However, the role of context in medicine cannot be minimized, as is underlined by visual classification results shown in (37), where inclusion of clinical parameters increases classification results by 7%.

A diagnosis needs to be made in the context of the clinical history of the patient. A similar concept was also already described for image retrieval in the non-medical domain, where the context of images in the text were used to improve visual image retrieval and vice versa (26). It cannot be performed in isolation based on just an image or series of images. The imaging modality, equipment, protocols and other factors of the image acquisition as well as age, gender, and clinical history of the patient can all impact the interpretation of an image. It would be difficult for humans as well as computer systems to try to diagnose with an image out of its clinical context.

We will review some examples from clinical practice where the role of context becomes apparent. Figure 1 presents CT images of two lungs, both of normal (healthy) patients. The image on the right is of an older patient which can resemble a diseased lung in a much younger patient. Here the context of age of the patient could potentially change the diagnosis from a pathological finding to a normal finding. We can see that the average density of the older patient’s lung is slightly higher as well, adding to the differences.

In Figure 2, the goal of the imaging study 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.

The display settings for lung tissue, bone, or soft tissue are different and the same image can look different depending on the acquisition and viewing conditions. In the image on the right, one can observe the texture of the lung tissue but other soft tissue or bones are not as easy to visualize while for the image on the left, the texture of the lung tissue cannot be discerned. The context of the goal of the imaging study is relevant in determining the pathology in the image as one would be unable
Figure 1. The two images show the significant changes in lung texture of a healthy patient of a different age, Figure (a) of a 25 year-old person and Figure (b) of an 88 year-old person.

![Image](a)

![Image](b)

Figure 2. Two CT scans of the lung shown in a varying level/window setting as the images were taken with a different goal in mind; image (a) was taken to analyse the mediastinum and image (b) to analyse the interstitial lung tissue. Although of the same modality and exactly same anatomic region comparing images taken with a differing goal in mind does often not make much sense.

![Image](a)

![Image](b)

to find diseased lung tissue in an image with acquisition or display goal being the imaging of the mediastinum.

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. Figure 2b shows the development of radiation fibrosis on a patient with radiation therapy.
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. Heart problems change the lung tissue particularly in lung CTs due to changes in blood flow and a resulting increased density in the tissue. Other contextual informations that need to be considered when retrieving images include changes in image acquisition techniques, equipment, resolution, contrast agents, and protocols.

THE FUTURE OF IMAGE MANAGEMENT IN CLINICAL PRACTICE

Multimodal approaches to image retrieval can be extremely useful as seen in the ImageCLEF experience (17,39). Some queries are better suited to visual techniques while others are best handled by textual methods. Clinical data is often incomplete, unstructured, and varied in levels of specificity and detail. Combining various data sources can be valuable in providing the context for these images. This can include the use of the free-text accompanying the images, structured data explaining the context of the image, textual descriptions of the image content, and electronic patient record, etc. Visual techniques need to be able to accommodate manual interventions for extractions from regions of interest and task specific segmentations as well as registration on a local level. Such toolboxes need to be made available to accommodate images from different acquisition systems and be extendible as imaging technology advances. More intuitive ways of formulating a query including the ability to upload multiple sample images of varying modalities, to convey negation, and to perform multiple levels of relevance feedback.

It is also very important to create proper datasets that also include clinical information and particularly pathology. Having datasets annotated with only simple modality, anatomy, and viewing angle as in (27) can be used to test algorithms and for fully automatic very low level tasks but can unfortunately not really help clinical applications. Users also state that pathology is the most important search criterion (24).

Such datasets need to be made available for a larger public to make sure that their knowl-
edge can be fully exploited (31). One way of doing so are the use of Web 2.0 techniques to create datasets and share medical knowledge (32,33). One system aiming at this is MDPixx, and creating data sets in this way may be much less costly than having a central organization for annotation image and marking regions of interest. Google (or other search engines) will be used for diagnosis in one way or another (34) whether we like this development or not. Another networking technology to take into account are grid networks (35) that could deliver the necessary computing power to treat full PACS archives and at the same time better use an existing IT infrastructure in medical institutions that often do not have research computing infrastructures in place.

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.

All this means leaving the comfort zone of retrieval of similar images to a single example image, a research domain that has been well explored. The result would be a case-based retrieval system that can integrate several images of the same or varying modalities, plus structured data and free text, linking a large variety of knowledge sources such as ontologies or external literature.

**CONCLUSION**

Management of medical images is becoming increasingly important as the number and variety of images being created in medical settings everyday is growing rapidly. Importance is diagnosis is equally increasing. Content-based image retrieval or techniques based on the query-by-example paradigm have been studied extensively in computer vision. However, the global, low level visual features automatically extracted by these algorithms do not always correspond to the concepts that a user has in his mind for searching. The role of images in diagnostic medicine can be complex, making an interpretation of the images hard for a medical doctor who might not be a specialist in all exams undertaken or all anatomic regions. Image retrieval can in these cases deliver important information to help interpret a given case or set of images by supplying similar other cases that might also be similar in diagnosis.

In this paper we state that purely visual techniques for medical image retrieval may not be sufficient for most clinical applications. In medicine, visual information taken alone, and thus out of its clinical context, is less meaningful than the images viewed in the context of the patient and the environment. We believe that purely visual CBIR methods in medicine have not lived up to expectations and seem only be suitable for very precise and simple applications such as turning lung x-rays into the right orientation (28), detecting modality (29), or for extract very simple concepts from medical images such as in the automatic image classification task of ImageCLEF (30).

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. These include multiple images, free text, structured data as well as external knowledge sources and ontologies. These 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 outcome) at the right time and in the right format.

**REFERENCES**

Bui AAT, Taira RK, Dionisio JDN, Aberle DR, El-Saden S, Kangarloo H. Evidence-based radiology:

Miller GA. The magical number seven plus or minus two: some limits on our capacity for processing information. Psychol Rev. 1956 Mar;63(2):81-97.


ENDNOTES

1 http://pubimage.hcuge.ch/
2 http://www.benchathlon.org/
3 http://www.imageclef.org/
4 http://www.imageval.org/
5 http://www.google.com/image/
6 http://www.yahoo.com/
7 http://goldminer.arrs.org/
8 http://www.mypacs.org/
9 http://mirc.rsna.org/
10 http://www.mdpixx.com/
Please Insert Bios