Multi-modal Content Based Image Retrieval in Healthcare: Current Applications and Future Challenges

Jinman Kim[1], Ashnil Kumar[1], Tom Weidong Cai[1] and David Dagan Feng[1,2]

1 Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, Australia

2 Center for Multimedia Signal Processing, Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong

Abstract

Modern healthcare environments have become increasingly reliant on medical imaging and this has resulted in an explosive growth in the number of imaging acquisitions obtained as part of patient management. The recent introduction of multi-modal imaging scanners has enabled unprecedented capabilities for patient diagnosis. With multi-modal imaging, two or more complementary imaging modalities are acquired either sequentially or simultaneously e.g. combined functional positron emission tomography (PET) and anatomical computed tomography (CT) imaging.

The efficient and accurate retrieval of relevant information from these ever-expanding patient data is one of the major challenges faced by applications that need to derive accumulated knowledge and information from these images, such as image-based diagnosis, image-guided surgery and patient progress monitoring (patient's response to treatment), physician training or education, and biomedical research. The retrieval of patient imaging data based on image features is a novel complement to text-based retrieval, and allows accumulated knowledge to be made available through searching. There has been significant growth in content-based image retrieval (CBIR) research and its clinical applications. However, current retrieval technologies are primarily designed for single-modal images and are limited when applied to multi-modal images, as they do not fully exploit the complementary information inherent in these data e.g. spatial localization of functional abnormalities from PET in relation to anatomical structures from CT.

Multi-modal imaging requires innovations in algorithms and methodologies in all areas of CBIR, including feature extraction and representation, indexing, similarity measurement, grouping of similar retrieval results, as well as user interaction. In this chapter, we will discuss the rise of multi-modal imaging in clinical practice. We will summarize some of our pioneering CBIR achievements working with these data, exemplified by a specific application domain of PET-CT. We will also discuss the future challenges in this significantly important emerging area.

Index Terms

Content-based image retrieval

Multi-modal image retrieval

Multi-modal imaging

PET-CT

Image segmentation

Image registration

Attributed relational graphs

Section 1. Introduction

Content-based image retrieval (CBIR) refers to the use of the visual attributes of images for searching an image database. In recent years, we have witnessed a rapid rise in CBIR research and the development of CBIR based clinical applications for medical image databases (Müller, 2004; Cai, 2007; Deserno, 2007; Long, 2009; Kim, 2009). Some well-known CBIR investigations include the retrieval of high-resolution lung computed tomography (CT) introduced by Shyu (1999); a study by El-Naqa (2004) for the retrieval of microcalcification types from mammography images; the retrieval of dynamic positron emission tomography (PET) images based on temporal attributes (Cai, 2000; Kim, 2006); and more recently, a retrieval system for spine X-ray images using a partial shape matching approach (Xu, 2008).

The aforementioned CBIR systems were designed for a single type of imaging modality, and were thus able to utilize domain specific knowledge and image processing optimizations. Such approaches, however, may be limited in their application when applied to different imaging modalities. There are several CBIR studies that are not bound to a single modality and that aim at supporting a diverse range of medical images. For example, in Lehmann (2005), an automatic categorization for a wide variety of medical images was presented that allowed for a robust classification of medical images. Their results demonstrated that their categorization technique, which based on global image textural features and scaling, was successful in classifying images according to their anatomical regions, imaging modality and specific orientation. The introduction of ImageCLEFmed, a medical section of the Cross Language Evaluation Forum (CLEF), has led to increasing interest in benchmarking the automatic classification and information retrieval from diverse medical image modalities (Deselaers, 2009; Rahman, 2007).

ImageCLEFmed has created a standard environment for the evaluation and improvement of medical CBIR from heterogeneous collections containing images as well as text information.

However, regardless of their ability to retrieve from multiple modality databases, current retrieval technologies are inherently designed for single-modal images. Thus, these algorithms and systems are limited when applied to multi-modal images, as they do not fully utilize the additional complementary information that may be derived from these images. In this chapter, we refer to multi-modal images as two or more medical image modalities that are co-aligned to each other. These separate modalities maybe co-aligned through sequential or simultaneous acquisition by a hybrid scanner or via image processing (see Section 2 for more details). Significant clinical benefits have arisen from the use of these multi-modality images and this has led to rapid acceptance of these images in clinical practice (Schulthess, 2009; Townsend, 2004). For example, the recently invented hybrid scanner that combines PET and magnetic resonance imaging (MRI) in a single scan (Beyer, 2009), enables the visualization of the functional abnormalities from PET (e.g. tumours) in relation to its co-aligned anatomical counterpart from MRI (soft and hard tissues) for the first time. These multi-modal images introduce new challenges and opportunities for CBIR research and development.

Apart from medical imaging, there has been great interest in multi-modal retrieval in consumer, public safety and professional applications (Kankanhalli, 2008). In these multimedia information retrieval (MIR) approaches, large array of modalities e.g. video (i.e. surveillance), text, signals and sound (i.e. voice recognition), in addition to image modalities (i.e. satellite), are combined for information fusion which are then used for information retrieval. Most common approach to

multi-modal information fusion is by combing the semantic information that is derived from text to complement and improve the image features that are automatically extracted. Such combination has shown success in enhancing the image representation for retrieval (Zhang, 2005; Fu, 2008). In Kumar (2010), object detection in dynamic environment was proposed where several complementary modalities like visible spectrum and thermal infrared video are fused using evidence theory. Such multi-modal techniques share many complementary techniques with multi-modal medical CBIR and their combination may lead to accelerate breakthrough in CBIR research.

In this chapter, we present the state-of-the-art in multi-modal CBIR in medical imaging domain and also the emerging research challenges. We briefly introduce recent advances in biomedical multi-modal imaging scanners in Section 2. An emphasis is placed on PET-CT imaging, which is the modality used in our multi-modal CBIR research. In Section 3, we summarize some of experimental retrieval results with PET-CT and in Section 4, we discuss the major challenges and future work for multi-modal CBIR.

Section 2. Multi-modal Biomedical Imaging

During the past decade, there has been rapid development in multi-modal scanners, which acquire two separate modalities sequentially, usually within a single examination. This has resulted in the production of co-aligned images, such as combined PET-CT. PET-CT has enabled functional information from PET to be assimilated with its anatomical counterpart in CT, thereby introducing new and improved diagnostic capabilities (Beyer, 2009). Figure 1 shows an example of a PET-CT scan for use in oncology. Single photon emission computed tomography (SPECT) combined with CT, the SPECT-CT, is another multi-modal scanner, as combined imaging has enabled a wider acceptance of SPECT as a quantitative imaging modality (Chowdhury, 2008).

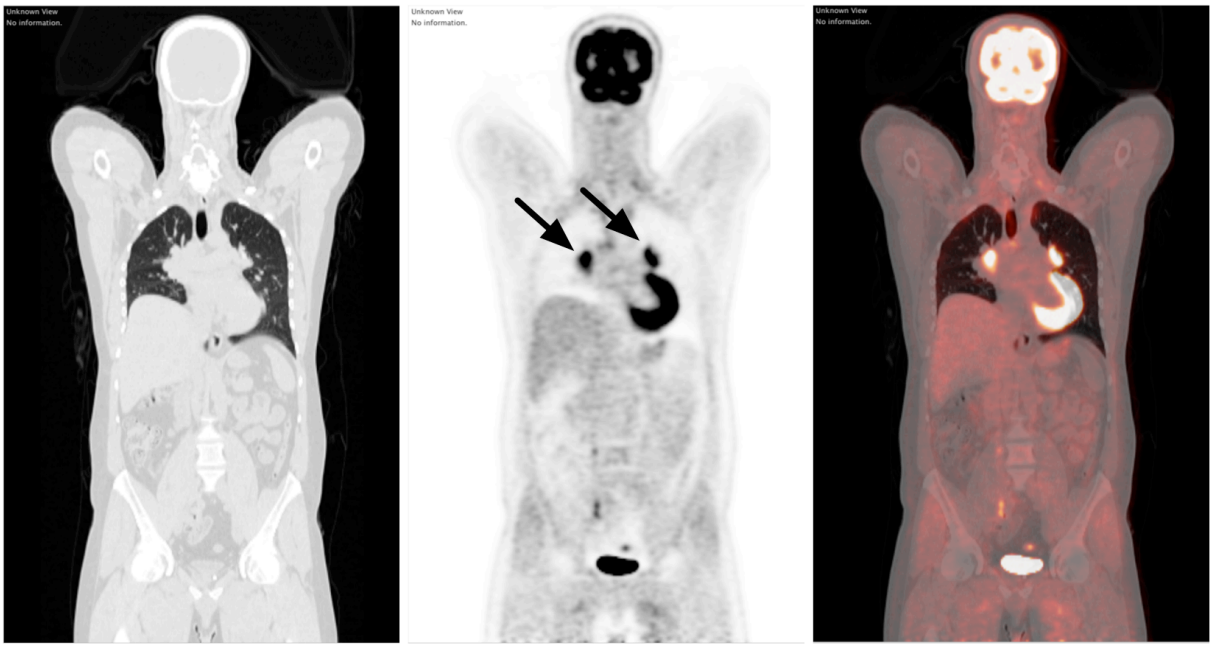

Figure 1. A PET-CT image of a patient diagnosed with lung cancer. The functional PET image (middle) clearly depicts the increased glucose evident in the lungs and its surrounding structures (indicated by arrows). The anatomical CT image (left) provides the sharp boundary of the surrounding structures. The fusion of the two images (right) shows the combined functional and anatomical information. The lookup tables (LUTs) used in the CT, PET and fusion images are grayscale, inverted grayscale and 'red hues on white' for PET fused with grayscale for CT, respectively.

The most promising new addition to multi-modal imaging is the multi-modal PET-MR scanner, a hardware combination that brings rich tissue definition (from the MRI) to functional PET images. Several new and different features are introduced by replacing CT with MRI. MRI is a high resolution anatomical imaging modality that offers better soft-tissue contrast resolution and a larger variety of tissue contrasts than CT. It also allows for the acquisition of functional MRI, thereby enabling the temporal correlation of blood flow with metabolism or receptor expression

in brain studies and, more importantly, allowing the assessment of flow, diffusion, perfusion, and cardiac motion within a single examination (Zaidi, 2007). Thus, PET-MR will be the imaging modality of choice in certain clinical cases, such as neurology and musculoskeletal applications.

These multi-modal scanners are already utilized in routine clinical practice (PET-CT and SPECT-CT) or are in the process of being used (PET-MR). However, there are many more multi-modal imaging technologies that are currently in research and development. As an example, ultrasound (US) with its ability to capture in real-time the dynamics of tissue vascular motion (typically in 2D) has benefited from the fusion of high resolution and volumetric anatomical imaging from MRI. Such an approach has found applications in the visualization of carotid arteries (Tang, 2007) and in research aimed at improving cardiac imaging and real-time respiration control (Feinberg, 2010). Another multi-modal imaging development is in the new field of neuro-imaging, where functional MRI is coupled with either electroencephalography (EEG) or magnetoencephalography (MEG) for potentially understanding neuronal activation and activation center communication and processing (Moseley, 2004).

Although multi-modal scanners provide hardware-based co-registered images, multiple modalities have been and will continue to be aligned using software-based image registration algorithms. Image registration has reached a high level of automation and robustness, and has resulted in many clinical investigations, e.g. the work by Maes (1997) on multi-modal image registration by maximization of mutual information. Such approaches also have the ability to fuse temporal data i.e. intra-patient data spreading over multiple datasets acquired for assessing

response to treatment; and for inter-patient registration i.e. building a statistical atlas for use in automated segmentation (Commowick, 2008).

Section 3. Multi-modal PET-CT CBIR

Our Biomedical and Multimedia Information Technology (BMIT) research group has been active in the research of novel CBIR solutions for PET-CT images since the introduction of multi-modal PET-CT scanners into routine clinical practice. PET-CT scans present significant advantages in patient diagnosis and management, but also place new challenges in computerized image analysis and its application to CBIR. Prior to PET-CT, our BMIT group has worked on CBIR for dynamic PET images in both 2D (Cai, 2000) and 3D (Kim, 2006) domains and these results, together with challenges in multi-dimensional medical CBIR were summarized in Kim (2009). To the best of our knowledge, we are the pioneers in proposing new CBIR methodologies for multi-modal PET-CT images.

Our main innovation in multi-modal CBIR is the use of complementary information that is derived from combining both the functional (from PET) and anatomical (from CT) images. We suggest that the exploitation of the unique multi-modal imaging attributes such as the complimentary information and the spatial relationships between the modalities can lead to new and improved approach to image retrieval. This section will present some of our research findings towards developing CBIR systems for PET-CT images. All our studies were evaluated by experimentation on clinical and simulated PET-CT databases of lung cancer patients.

Section 3.1. Combined PET-CT Feature Extraction

Our initial study in Kim (2007a) investigated the use of multi-modal features extracted from the PET-CT images. In this study, our novelty was in using automated multi-modal segmentation algorithm to extract regions of interest (ROIs) that comprises of complementary features from

PET (e.g. tumours and other abnormal regions) and CT (e.g. structural definition of the lung and surrounding structures). We derived image features including shape, size, and average pixel values from the segmented ROIs. These features were then used for measuring the similarity the PET-CT data sets. We tested our method with a PET-CT lung cancer database and examined its ability to retrieve cancer patients that shared similarities to an input image. The search was based on a visual query sample e.g. an image that consists of a tumour of size *x* pixels residing in the right lung. The similarity of tumours was calculated based on the size (pixel count) difference between the PET query images and the PET images in the database. The retrieved PET features were then used to check if the tumour belonged in the left or right lung according to its corresponding CT features. Our preliminary results on 10 controlled patient data (5 lung cancer and 5 healthy patients) indicated that the dual-modal feature extraction enabled a new approach to PET-CT CBIR. This study however didn't exploit the full range of spatial attributes between the modalities and also was limited in its searching capabilities.

Section 3.2. Graph-based PET-CT Retrieval

The use of spatial relationships for CBIR of dual-modal images was inspired by the fact that the interpretation of biomedical knowledge relies on knowing where structures are located in relation to each other. In particular, the extent to which pathology-bearing regions are involved with anatomical structures is vital for diagnosis and treatment planning. In the context of cancer imaging, the old (Mountain, 2000) and new (Detterbeck, 2009) TNM staging systems for lung cancer, and the Ann Arbor staging system (Carbone, 1971), specify the stage of the disease according to the spatial location of tumours in relation to anatomical structures.

Several studies have demonstrated the use of disease location as features for CBIR. The retrieval of lung disease images was examined in (Shyu, 1999; Aisen 2003). In these studies the feature set included pathology-bearing regions and the lung lobular regions in which they occurred, in addition to traditional texture and shape features. In Petrakis (2002) a number of different spatial similarity approaches were evaluated for the retrieval of synthetic MRI images. The study discovered that graph-based techniques, while being the least computationally efficient, were the most accurate. However, all of these techniques only considered single modality images.

In Kumar (2008; 2009) we proposed a graph-based retrieval methodology that was able to preserve important spatial relationships between the multi-modal images. In our method, every ROI forms a graph vertex, and spatial relationships between ROI are represented by edges between vertices in an Attributed Relational Graph (ARG). Our novelty lies in the indexing of inter-modality relationship features i.e. attributes on an edge that is incident CT vertex and a PET vertex. Figure 2 depicts our graph construction process and highlights the extraction of the inter-modality feature calculation. Figure 2(a) shows the creation of the vertices representing CT ROI, while Figure 2(b) depicts the creation of the PET vertex representing the tumour. Finally, the graph (in Figure 2(c)) was constructed after the calculation of relationship features.

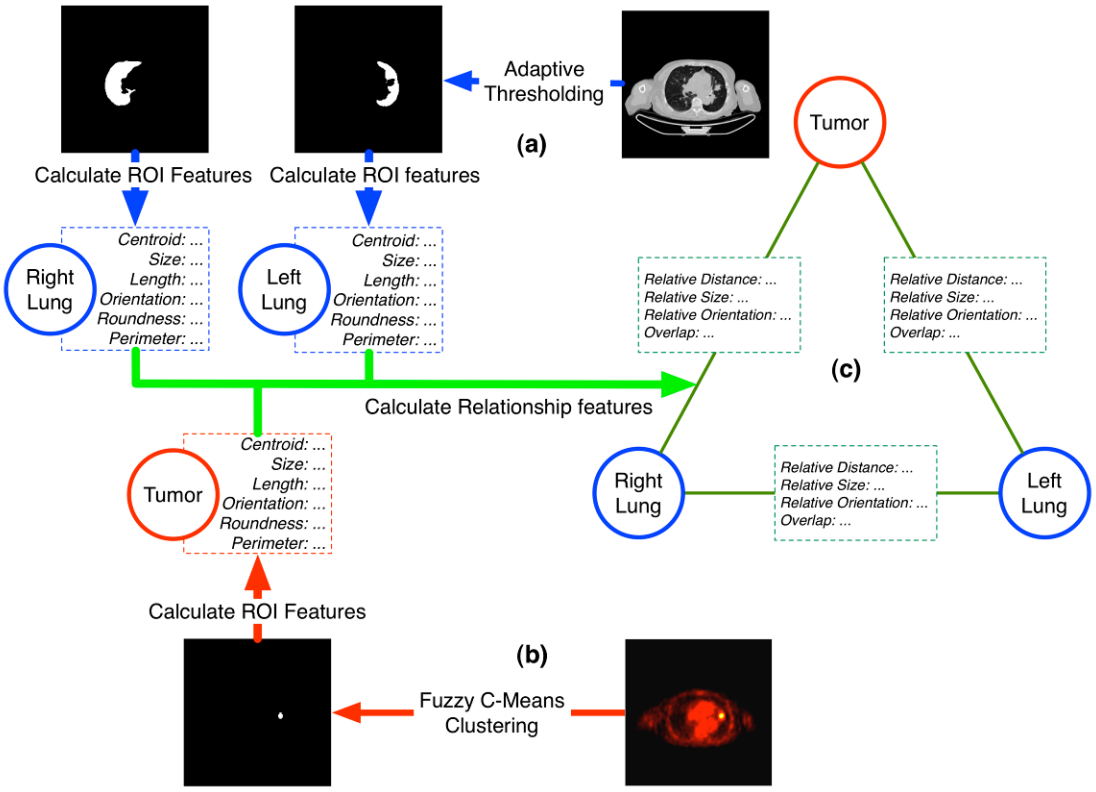

Figure 2. Graph construction process for multi-modal PET-CT datasets. The CT (a) and PET (b) vertices index the features for their ROI and are used to construct the ARG (c), which preserves all the features and the relationships between the vertices.

In our studies, we define our dual-modality graph $G = (V, E)$, where V is the set of graph vertices and E is the set of graph edges. Every vertex is a feature vector of the form $v_i = <C, a, p, o, r, l>$, where $C = (c_x, c_y)$ is the centroid of the ROI, and a, p, o, and r are the area, perimeter, orientation and roundness of the ROI, respectively. The maximum internal length between two pixels on the boundary of the ROI is given by l. Every edge e_{ij} between two vertices v_i and v_j, where $i \neq j$, is a feature vector of the form $<rd, ro, ov, ra>$ where rd is the distance between the centroids of the ROI, ro is the angle between the centroids of the ROI, ov is the overlap between the two ROI, and ra is the ratio of the areas of the two ROI. The combined vertex and edge feature vectors

allow us to represent complex semantic relationships, such as "the tumour of size *x* lies within the lower half of the left lung, near the lung boundary".

From the PET-CT images, the ROIs were selected by segmenting the CT images using adaptive thresholding with refinements (Hu, 2001). Tumour ROIs were selected from the PET images via fuzzy c-means clustering (Kim, 2007b). The centroid feature was used primarily for the calculation of *rd* and *ro*. It was not used as part of the similarity measurement that we describe later. The features indexed as attributes were normalized using the procedure described in (Petrakis, 2002), to achieve scale, translational and rotational invariance. We examined only spatial features for these ROI, similar to the approach in (Petrakis, 2002). However, our graph model is still capable of indexing almost any feature as a graph attribute.

Section 3.2.1. Graph Similarity Matching

The multi-modal image database from which similar images are retrieved have their graphs constructed as an offline process. These graphs are stored in an index for retrieval. Image retrieval is achieved by comparing the graph of a query image, to the graphs stored in the index. Essentially, this is a subgraph matching problem, where the graph attributes are used to determine the degree of similarity between subgraphs with the same structure. As this task is *NP-complete*, we limit our work to small graphs representing only part of the whole body e.g. the lungs.

The matching process between a query graph, $G_Q = (V_Q, E_Q)$, and an indexed graph, $G_S = (V_S, E_S)$, attempts to find structure preserving mappings between vertices and edges on both graphs.

We are attempting to find all isomorphisms $\phi: G_Q \rightarrow G_S$, which satisfy the following condition based on its vertex and edge mapping functions, $\phi_v: V_Q \rightarrow V_S$ and $\phi_e: E_Q \rightarrow E_S$, respectively. If $\phi_v(v_{Q1}) = v_{S1}$ and $\phi_v(v_{Q2}) = v_{S2}$, where $v_{Q1}, v_{Q2} \in V_Q$ and $v_{S1}, v_{S2} \in V_S$, then $\phi_e(e_{Qa}) = e_{Sa}$ where $e_{Qa} \in E_Q$ is the edge incident to v_{Q1} and v_{Q2}, and $e_{Sa} \in E_S$ is the edge incident to v_{S1} and v_{S2}. Often multiple structure-preserving isomorphisms between two graphs will be discovered because the mapping works purely on graph structures without any reference to the clinical ROI that each vertex represents. We deal with this problem by selecting the best isomorphism through comparing the feature attributes in each mapped graph element. We define the best isomorphism as the isomorphism in which the sum of attribute differences of mapped graph structures is minimum. As such, we apply the Euclidean distance on each individual vertex and edge mapping:

$$d(q,s) = \sqrt{\sum_{i}^{N} \left(q_i - s_i\right)^2} \qquad (1)$$

where q is a vertex or edge of Q, s is a vertex or edge of S and $\phi(q) = s$, N is the number of vertex or edge attributes, q_i is an attribute of the query vertex or edge and s_i is an attribute of the mapped vertex or edge. The distance represents the feature difference between mapped vertices and edges. When we sum the differences of all the mappings in an isomorphism, we get a total cost for the isomorphism. The best isomorphism results in the lowest cost from G_Q to G_S.

When applied across index, this process finds from every indexed graph the best isomorphic subgraph and the cost for the matching. We use the costs to rank the images represented by the indexed graphs. Lower costs represent a greater degree of similarity between the query graphs and the database graphs, and thus a greater similarity between the query image and the database

image represented by the graph. We rank the database images in ascending order based on the cost to match their graph to the query image's graph.

Section 3.2.2. Experimentation and Results

In (Kumar, 2008), we created a set of over 100 simulated PET-CT images by applying a series of controlled random variations to a set of segmented clinical PET-CT images. We searched this simulated data set using a number of different queries: lungs with no tumours, lungs containing a single tumour and images with two tumours. In all cases our method was able to distinguish images that were spatially similar. Furthermore, we also used one of the template images that were used to create the simulation database as the query. We discovered that in this case, the simulated variations derived from the template in question achieved higher rankings than those made from other templates. In this way we showed how our proposed technique could be used for exact and inexact PET-CT image retrieval.

We also showed the clinical potential of our retrieval methodology by carrying out a series of searches on a data set of 21 clinical PET-CT 2D slices, from 10 patient studies. We used a query-by-sketch approach wherein users drew the shape of the lungs and placed tumours within them. This sketched image was then processed as a query and the most similar images were retrieved. Figure 3 shows the query and retrieved results of a search on the clinical data set. A user who wished to locate images with at least one tumour in the left lung (the left lung appears on the right side in CT images) created the query image. The retrieved results all have the property the user desired. The best result has a single tumour of a similar size to that drawn by the user, the second best result has two tumours smaller than those specified by the user and the third result

has tumours that are even smaller, but in the same spatial location as the tumour in the query sketch.

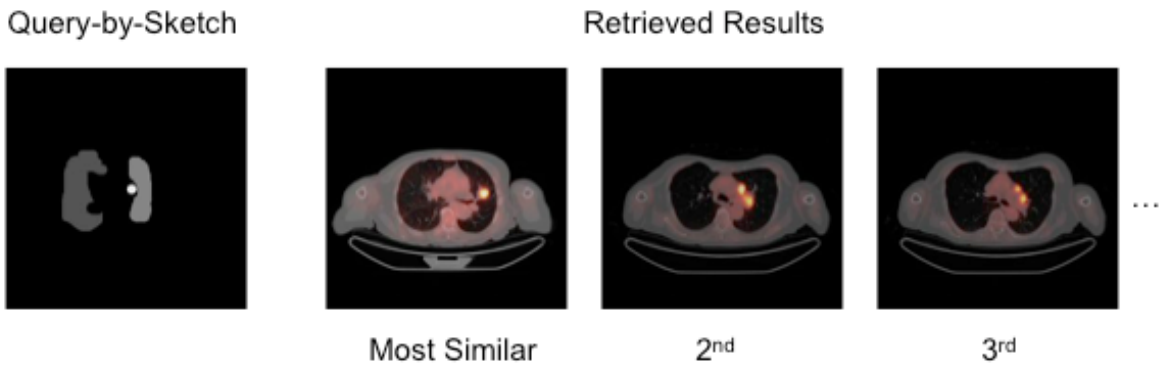

Figure 3. Example retrieval results (Kumar, 2008). Compared to the query image, the result has correctly identified the image with only a single tumour as the best match.

Furthermore, in (Kumar, 2009) we investigated the retrieval in the specific cases where tumours invaded out of the lungs and into the chest wall. More specifically, we explored whether the retrieval technique could differentiate between images where chest wall invasion was occurring and those images where the tumours were located near the pleural surface. For this experiment, we expanded our simulated database to over 400 PET-CT images, half of which contained chest wall invasion. The simulated tumors were created using an approach similar to that in (Korn, 1998). We examined the 25 highest ranked retrieved results in a number of different searches for images where the tumour invaded the chest wall. We found that on average 76% of the retrieved images contained chest wall invasion. In order to increase the accuracy of our system, we introduced weighting to our similarity measurement to Equation 1:

$$d(q,s) = \sqrt{\sum_{i}^{N} w_i \left(q_i - s_i\right)^2} \qquad (2)$$

where w_i is the weight for a particular vertex or edge attribute. All other elements are defined as in Equation 1. Using empirically derived weights, we discovered that 88% of the retrieved images contained chest wall invasion.

Section 3.3. Grouping Similar PET-CT Retrieval Results

Our technique is capable of retrieving images that are only slight variations of the query, as demonstrated by the results of Kumar (2008, 2009). However, this means that when the dataset contains images of similar attributes from patients who have had no change in their condition (neither disease progression nor any response to treatment), then it is expected that these similar images will form the majority of the retrieved results. In CBIR, the motivation is to retrieve similar but distinct results, based on the user specified attributes. For example, retrieval-based diagnostic systems such as in Aisen (2003) rely upon a variety of similar but distinct medical images being retrieved for interpretation by physicians. However, retrieval algorithms based on conventional similarity measures would typically return results that are very similar to the query, especially with large database samples (e.g. in excess of several thousands), thus reducing the information to only a few similar cases that have almost no distinction between them. Therefore, we proposed a grouping technique for very similar results as a means of obtaining a greater number of distinct cases, thereby providing improved retrieval diversity in very large patient image databases.

Our grouping technique is inspired by the near-duplicate detection method proposed by Wang (2007) for filtering image spam. We have integrated a near-duplicate detector which groups the retrieved results based on its relative similarity. In this approach, similar results without

significant variations (near-duplicates) can be grouped together and thus increasing the diversity of the results. Our near-duplicate detector calculated the Manhattan distance on the Haar coefficients of the ROIs corresponding to the mapped vertices. Where the distance was *below* a set threshold, the results were marked as near-duplicates. The results so marked were then grouped allowing other more distinct images to be retrieved.

We evaluated our approach on the graph-based CBIR system (described in Section 3.2) using the proposed grouping described above. To simulate a large database, we generated a data set of over 6000 simulated PET-CT images with every query having at least five associated near-duplicate images in the data set. Our experimental procedure involved executing 24 different queries. The images used as queries varied in lung shape, disease location and tumour size. They were constructed to allow us to test retrieval and near-duplicate detection of images with tumours in specific regions, or having certain properties e.g. tumour size. These queries were invoked to examine the effectiveness of the detector, its false positive rate, and the precision of the retrieval in the top 10, 15 and 20 results without and with the near-duplicate grouping. The results for the tests were averaged. In addition, we also examined the number of near-duplicate images that were successfully detected. Table 1 summarizes the results of our experiments with a query image consisting of a single tumour within a lung.

Table 1. Retrieval results for near-duplicate filtering.

Average Precision (%)	Pre-Grouping	Post-Grouping
Top 10 results	92.50	84.17
Top 15 results	88.89	82.22
Top 20 results	86.67	79.58
Detection Effectiveness over all Experiments		
Total near-duplicates	180	
Near-duplicates detected	168	
False positives (non-duplicates)	13	
Percentage effectiveness	93.33	
False positive rate	7.2%	

We were able to detect 93% of all near-duplicates, as measured across all the experiments. The near-duplicate images that were not detected all contained tumours that substantial difference in size when compared to the query image i.e. the false negative results contained significant (about 200%) changes in only a subset of the ROIs. Better performance in detecting near-duplicate images was measured when smaller variations were spread over all the ROI.

In an CBIR application, it is important to ensure that the near-duplicate grouping does not degrade the retrieval results by having too high a false positive rate. The Haar wavelet filter was chosen because of its low false positive rate in prior studies (Wang, 2007). In our experiments, the false positive rate did not rise above 7.2%, meaning that very few images were incorrectly detected as near-duplicates. Our results in Table 1 show that there is a drop of 7% to 9% in precision in the post-detection and grouping results compared to the when there is no near-duplicate grouping. Our simulation was designed so that a number of near-duplicate results would be clustered near the highest rankings, and as such, the precision decrease due to their

removal is expected. Despite this drop, our proposed retrieval system was able to maintain high levels of precision.

Section 3.4. Multi-modal Retrieval without the Reliance on Well-defined Feature Sets

In the studies above, we relied on robust and accurate segmentation results. Although there have been significant advances in image segmentation (Hu, 1991; Commowicka, 2008; Korfiatis, 2008), the ability to minimize dependence on segmentation is important. It enables greater automation for feature extraction and could potentially lead to reduced degradation of the retrieval results caused by segmentation errors. In addition, the speed of execution can also be improved by the removal of segmentation computational requirements. In our recent study (Song, 2010), we investigated an approach to PET-CT retrieval that did not rely upon accurate segmentation results. In this study, visual patterns were represented using texture features extracted with Gabor filtering, and further transformed into discrete pattern categories. Some incorrect categorizations resulted from the inclusion of surrounding tissues during the lung field estimation, and a lack of sufficient differentiating information in the feature vectors. In order to reduce the occurrences of mismatching, we refined the categories with two binary SVMs, which have been shown to be an effective approach to lung CT classification (Korfiatis, 2008). Similarity measures were performed on the signature distribution bins with tunable weightings. The retrieval experiment was conducted on 870 clinical PET-CT image pairs selected from 20 patient studies with various stages of lung cancers. In our preliminary results, as a measure of precision, we counted the successful retrieval of top four and eight results. With four results, ~ 81.4% of the retrieved images exhibited the same visual texture patterns as the query image; this was lowered to 75.3% with eight results.

Section 4. Discussions and Future Work

The importance of CBIR in biomedical imaging is clear. Over the past decade, we have witnessed new innovations in CBIR that has led to the development of automated image categorization, evidence based diagnosis, computer aided diagnosis, retrieval for education and training, as well as towards developing classified biomedical image repositories. The recent trend in the development of biomedical imaging scanners has been towards multi-modal acquisitions and has introduced exciting new challenges in CBIR research. In order to maximize the retrieval potential of these multi-modal images, new algorithms that harness the inherent relationships between these images most be developed. The previous sections presented some of our multi-modal PET-CT research. Although we anticipate that our research will find wide applications in other multi-modal retrieval studies, our research is only addressing a small subset of multi-modal retrieval problems and challenges. With the transition to multi-modal gaining ground, we anticipate an array of new studies that will accelerate the research on multi-modal CBIR in numerous medical applications.

In our retrieval evaluations, only a small set of controlled patient data were used to demonstrate the capabilities of our algorithms. Although useful, these small sets are not a complete representation of the real clinical database, which generally contain larger variations. As such, we also employed simulation datasets that were derived from clinical data. The simulation enabled a wider range of experiments to test our retrieval algorithms. Our current simulation is in 2D and we are working towards a 3D model for greater realism, thus enabling the simulation of the 3D spatial relationships between the image ROIs. In order to move to a larger clinical database, we also need to investigate the robustness of the automated feature extraction

algorithms used. Our study in Song (2010) suggests an approach to feature extraction that has a lower reliance on image segmentation. Although the initial results were encouraging, spatial relationships between the extracted features were not considered to be important; in this study we assumed that images with abnormal nodules at different locations are similar. Further extension of this work could incorporate the graph representation in Section 3.2 to include the spatial relationships from the use of graphs in addition to the advantages that arise from a lower reliance on difficult image segmentation.

We are currently investigating alternative approaches that will use established segmentation algorithms that have been proven for clinical studies, i.e. PET image thresholding (Vauclin, 2009) and incorporating error tolerance to our ARG to compensate for segmentation errors. Nevertheless, there has been remarkable progress in multi-modal image segmentation which may translate to advanced feature extraction for CBIR. In Gribben (2009), the combined use of multi-modal information was shown to improve PET-CT lung tumour segmentation. In another study, Avazpour (2009) presented PET-CT segmentation that used information from both the modalities towards extracting brochogenic carcinoma structures.

While CBIR has shown some success, one of the factors that limit its wider use is lack of user interfaces (UIs) that can simplify and provide intuitive interactions to the myriad information in large databases. As identified in the study by Deserno (2007), the “usability gap” is only weakly addressed and that only a relatively small fraction of the research effort is directed to addressing it. Furthermore, existing UIs are typically designed for modality specific systems and to work for only a single type of image (Heesch, 2008; Long, 2009). In multi-modal CBIR, this problem is compounded. The combined information from the two imaging modalities creates supplementary

data, some of which needs to be derived from the very images they assist in interpreting e.g. fused images and LUT adjustments for viewing wide dynamic ranges. The visualization of all this information by a retrieval system will quickly become overwhelming. This problem will only grow more potent as image quality improves and as the number of image acquisitions increase. In our retrieval systems, we have adopted a simple conventional UI design that shows the most relevant retrieval results as depicted in Figure 4. Although the UI is practical, a lot of user interactions are necessary to browse through all the information. We are currently investigating more efficient retrieval browsing techniques that will improve the user's interaction with the multi-modal retrieved results.

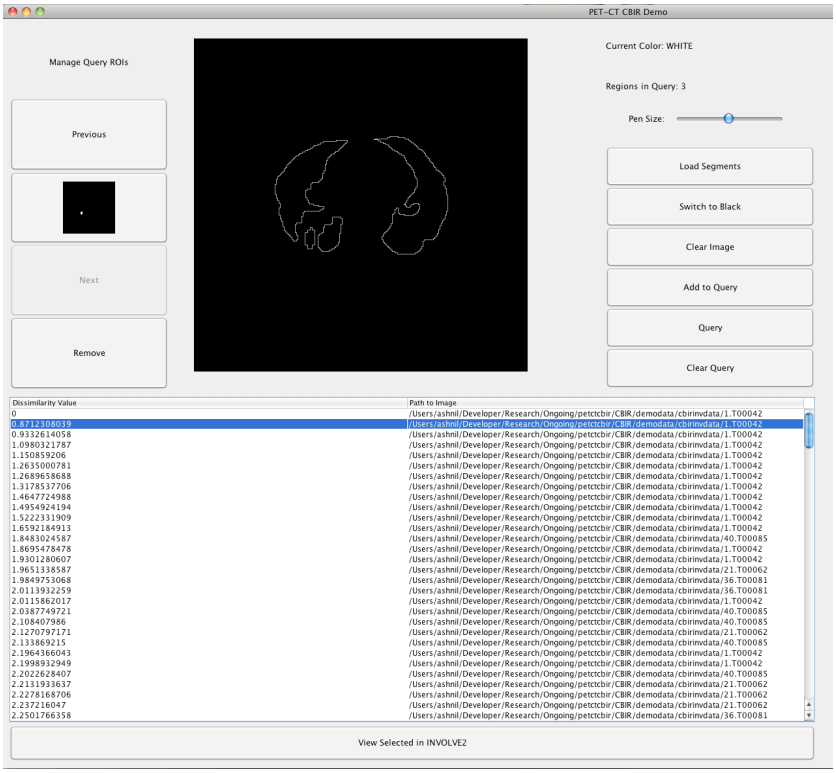

Figure 4. The user interface for our multi-modal PET-CT CBIR system. A query drawing interface is provided on the top. The query results populate the ranked list in the bottom with the similarity value and the path to the images. The image path may be selected which will then load a PET-CT image viewer.

Acknowledgement

We would like to thank the staff at the Royal Prince Alfred (RPA) Hospital of Sydney for their valuable contribution to our research and for providing the images for our experiments. This work was supported in part by ARC grants.

References

Aisen, A.M. Broderick, L.S. Winer-Muram, H. Brodley, C.E. Kak, A.C. Pavlopoulou, C. Dy, J. Shyu, C.-R. & Marchiori, A. (2003). Automated storage and retrieval of thin-section CT images to assist diagnosis: system description and preliminary assessment. Radiology. 228: 265-270.

Avazpour, I. Roslan, R.E. Bayat, P. Saripan, M.I. Nordin, A.J. & Abdullah, R.S.A.J. (2009). Segmenting CT images of bronchogenic carcinoma with bone metastases using PET intensity markers approach. Radiol Oncol. 43(3): 180-186.

Beyer, T. & Pichler, B. (2009). A decade of combined imaging: from a PET attached to a CT to a PET inside an MR. Eur J Nucl Med Mol Imaging. 36 (Suppl 1): S1–S2.

Cai, W. Feng, D. & Fulton R. (2000). Content-based retrieval of dynamic PET functional images. IEEE Trans Info Tech Biomed. 4(2): 152-158.

Cai, W. Kim, J. & Feng, D. (2007). Content-based medical image retrieval. In D. Feng (Ed.), Biomedical Information Technology (pp. 83-113), San Diago: Elsevier Press.

Carbone, P.P. Kaplan, H.S. Musshoff, K. Smithers, D.W. & Tubiana, M. (1971). Report of the committee on Hodgkin's disease staging classification. Cancer Research. 31: 1860-1861.

Chowdhury, F. Scarsbrook, A. (2008). The role of hybrid SPECT-CT in oncology: current and emerging clinical applications. Clin Radiol. 63:241–51.

Commowicka, O. Grégoirec, V. & Malandaina, G. (2008). Atlas-based delineation of lymph node levels in head and neck computed tomography images. Radiotherapy and Oncology

87(2): 281-289.

Deselaers, T. & Deserno, T.M. (2009). Medical image annotation in ImageCLEF 2008. Evaluating Systems for Multilingual and Multimodal Information Access. LNCS 5706: 523-530.

Deserno, T.M. Antani, S. & Long, L.R. (2007). Ontology of gaps in content-based image retrieval. J Digital Imaging. 22(2): 202-215.

Detterbeck, F.C. (2009). The new cancer staging system. Chest. 136: 260-71.

El-Naqa, I. Yang, Y. Galatsanos, N.P. Nishikawa, R.M. & Wernick, M.N. (2004). A similarity learning approach to content-based image retrieval: application to digital mammography. IEEE Trans Med Imag. 23:1233-1244.

Feinberg, D.A. Giese, D. Bongers, D.A. Ramanna, S. Zaitsev, M. Markl, M. & Günther, M. (2010). Hybrid ultrasound MRI for improved cardiac imaging and real-time respiration control. Magnetic Resonance in Medicine. 63(2): 290–6.

Gribben, H. Miller, P. Hanna, G.G. Carson, K.J. & Hounsell, A.R. (2009). MAP-MRF segmentation of lung tumours in PET/CT images. Proc IEEE ISBI. 290 – 293.

Heesch, D. (2008). A survey of browsing models for content based image retrieval. Multimed Tools Appl. 40: 261-284.

Fu, H.C. Xu, Y.Y. Pao, H.T. (2008). Multimodal search for effective image retrieval. Proc IEEE Conf IWSSIP 233-236.

Hu, S. Hoffman, E. A. & Reinhardt, J. M. (2001). Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans Med Imag. 20: 490-498.

Kim, J. Cai, W. Feng, D. & Wu, H. (2006). A new way for multidimensional medical data management: volume of interest (VOI)-based retrieval of medical images with visual and functional features. IEEE Trans Info Tech Biomed. 10(3): 598-607.

Kim, J. Cai, W. Feng, D. (2009). Bridging the feature gaps for retrieval of multi-dimensional images. International Journal of Healthcare Information Systems and Informatics. 4(1): 34-46.

Kim, J. Constantinescu, L. Cai, W. & Feng, D. (2007a). Content-based dual-modality biomedical data retrieval using co-aligned functional and anatomical features. MICCAI workshop on Content-based Image Retrieval for Biomedical Image Archives. 45-52.

Kim, J. Wen, L. Eberl, S. Fulton, R. & Feng, D. (2007b). Use of anatomical priors in the segmentation of PET lung tumour images. Proc IEEE Med Imag Conf. 4242-4245.

Korfiatis, P. Kalogeropoulou, C. Karahaliou, A. & Kazantzi, A. (2008). Texture classification-based segmentation of lung affected by interstitial pneumonia in high-resolution CT. Medical Physics. 35(12): 5290–5302.

Korn, P. Sidiropoulos, N. Faloutsos, C. Siegel, E. & Protopapas, Z. (1998). Fast and effective retrieval of medical tumour shapes. IEEE Trans Knowl Data Engineering. 10: 889-904.

Kumar, A. Kim, J. Cai, W. Eberl, S. & Feng, D. (2008). A graph-based approach to the retrieval of dual-modality biomedical images using spatial relationships. IEEE Conf EMBS. 390-393.

Kumar, A. Kim, J. Eberl, S. Fulham, M. & Feng, D. (2009). Content-based retrieval of PET-CT lung cancer data using spatial features. J Nucl Med. 50(Suppl 2): 176.

Kumar, P. Mittal A. & Kumar. P. (2010). Addressing uncertainty in multi-modal fusion for improved object detection in dynamic environment. Information Fusion. 11(4): 311-324

Lehmann, T.M. Guld, M.O. Deselaers, T. Keysers, D. Schubert, H. Spitzer, K. Ney, H. & Wein, B.B. (2005). Automatic categorization of medical images for content-based retrieval and data mining. Elsevier Computerized Medical Imaging and Graphics. 29: 143-155.

Liu, H. Zagorac, S. Uren, V. Song D. and Ruger, S. (2009). Enabling effective user interactions in content-based image retrieval. LNCS AIRS. 265-276.

Long, L.R. Antani, S. Deserno, T.M. & Thoma, G.R. (2009). Content-Based Image Retrieval in Medicine Retrospective Assessment, State of the Art, and Future Directions. Int J Healthc Inf Syst Inform, 4(1): 1–16.

Maes, F. Collignon, A. Vandermeulen, D. Marchal, G. & Suetens, P. Multimodality image registration by maximization of mutual information. IEEE Trans Med Imag. 16(2): 187-198.

Kankanhalli, M.S. & Rui, Y. (2008). Application potential of multimedia information retrieval. Proceedings of the IEEE. 96(4): 712-720.

Moseley, M. & Donna, G. (2004). Multimodal imaging introduction. Stroke. 35(Suppl 1): 2632-2634.

Mountain, C. F. (2000). The international system for staging lung cancer. Seminars in Surgical Oncology. 18: 106-115.

Müller, H. Michoux, N. Bandon, D. & Geissbuhler, A. (2004). A review of content-based image retrieval system in medical applications – clinical benefits and future directions. Int. J. Med. Informatics, 73: 1-23.

Petrakis, E.G.M. (2002). Design and evaluation of spatial similarity approaches for image retrieval. Image and Vision Computing. 20: 59-76.

Rahman, M. Bhattacharya, P. & Desai, B.C. (2007). A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. IEEE Trans Info Tech Biomed. 11(1): 58-69.

Schulthess, G.K. & Schlemmer, H-P.W. (2009). A look ahead: PET/MR versus PET-CT. Eur J Nucl Med Mol Imaging 36(Suppl 1): S3–S9.

Shyu, C. Brodley, C. Kak, A. Kosaka, A. Aisen, A. & Broderick, L. (1999). ASSERT: A physician-in-loop content-based image retrieval system for HRCT image databases. Compt Vision and Image Understanding, 75: 111-132.

Siadat, M-R., Soltanian-Zadeh, H., Fotouhi, F., & Elisevich, K. (2005). Content-based image database system for epilepsy. Elsevier Computer Methods and Programs in Biomedicine, 79, 209-226.

Song, Y. Cai, W. Eberl, S. Fulham, M. & Feng, D. (2010). A content-based image retrieval framework for multi-modality lung images", Proc IEEE CBMS. 285-290

Tang, A.M. Kacher, D.F. Lam, E.Y. Brodsky, M. Jolesz, F.A. & Yang, E.S. (2007). Multi-modal imaging: simultaneous MRI and Ultrasound imaging for carotid arteries visualization. Proc IEEE EMBS. 2603-2606.

Townsend, D.W. Carney, J.P. Yap, J.T. & Hall, N.C. (2004). PET/CT today and tomorrow. J Nucl Med. 45(Suppl 1): 4S-14S.

Vauclin, S. Doyeux, K. Hapdey, S. Edet-Sanson, A. Vera, P. & Gardin, I. (2009). Development of a generic thresholding algorithm for the delineation of 18FDG-PET-positive tissue: application to the comparison of three thresholding models. Phys Med Biol 54: 6901-6916.

Wang, Z. Josephson, W. Lv, Q. Charikar, M. & Li, K. (2007). Filtering image spam with near-duplicate detection. Proc Conf Email and Anti-Spam.

Xu, X.Q. Lee, D.J. Antani, S.K. & Long, L.R. (2008). A spine X-ray image retrieval system using partial shape matching. IEEE Trans Info Tech Biomed. 12(1): 100-108.

Zaidi, H. Mawlawi, O. & Orto, C.G. (2007). Simultaneous PET/MR will replace PET/CT as the molecular multimodality imaging platform of choice. Med Phys. 34: 1525-28.

Zhang, R. Zhang, Z. Li, M. Ma, W-Y. Zhang, H-J. (2005). A probabilistic semantic model for image annotation and multimodal image retrieval. Proc IEEE ICCV. 1: 846-851.