Medical Image Analysis 18 (2014) 330–342

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

A graph-based approach for the retrieval of multi-modality medical images

Ashnil Kumar [a,*], Jinman Kim [a], Lingfeng Wen [a,b], Michael Fulham [b,c], Dagan Feng [a,d]

[a] Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, Sydney, Australia
[b] Department of Molecular Imaging, Royal Prince Alfred Hospital, Sydney, Australia
[c] Sydney Medical School, University of Sydney, Sydney, Australia
[d] Med-X Research Institute, Shanghai Jiao Tong University, China

ARTICLE INFO

Article history:
Received 18 October 2012
Received in revised form 25 November 2013
Accepted 27 November 2013
Available online 6 December 2013

Keywords:
Content-based image retrieval
Graph similarity
Multi-modality
PET-CT

ABSTRACT

In this paper, we address the retrieval of multi-modality medical volumes, which consist of two different imaging modalities, acquired sequentially, from the same scanner. One such example, positron emission tomography and computed tomography (PET-CT), provides physicians with complementary functional and anatomical features as well as spatial relationships and has led to improved cancer diagnosis, localisation, and staging.

The challenge of multi-modality volume retrieval for cancer patients lies in representing the complementary geometric and topologic attributes between tumours and organs. These attributes and relationships, which are used for tumour staging and classification, can be formulated as a graph. It has been demonstrated that graph-based methods have high accuracy for retrieval by spatial similarity. However, naïvely representing all relationships on a complete graph obscures the structure of the tumour-anatomy relationships.

We propose a new graph structure derived from complete graphs that structurally constrains the edges connected to tumour vertices based upon the spatial proximity of tumours and organs. This enables retrieval on the basis of tumour localisation. We also present a similarity matching algorithm that accounts for different feature sets for graph elements from different imaging modalities. Our method emphasises the relationships between a tumour and related organs, while still modelling patient-specific anatomical variations. Constraining tumours to related anatomical structures improves the discrimination potential of graphs, making it easier to retrieve similar images based on tumour location.

We evaluated our retrieval methodology on a dataset of clinical PET-CT volumes. Our results showed that our method enabled the retrieval of multi-modality images using spatial features. Our graph-based retrieval algorithm achieved a higher precision than several other retrieval techniques: gray-level histograms as well as state-of-the-art methods such as visual words using the scale- invariant feature transform (SIFT) and relational matrices representing the spatial arrangements of objects.

1. Introduction

Medical imaging is fundamental to modern healthcare and its widespread use has resulted in the rapid expansion of image storage in databases and a variety of archival systems, e.g. picture archiving and communication systems (PACS). These image repositories store vital patient diagnostic imaging data and they offer the opportunity for evidence-based diagnosis, teaching, and research (Müller et al., 2004; Müller et al., 2007; Müller et al., 2009; Akgül et al., 2011; Müller and Deserno, 2011). These additional applications require methods to extract relevant features from the clinical images, index them in a representational structure, compare the similarity of representations, and present the comparative data back to a user.

Content-based image retrieval (CBIR) is an image search technique that complements the conventional text-based retrieval of images by using visual features as search criteria (Smeulders et al., 2000). These features typically include shape, colour, texture, and the spatial arrangement of objects within an image, as well as state-of-the-art image features such as the scale-invariant feature transform (SIFT) (Lowe, 2004), a powerful descriptor for object and image classification. Medical CBIR, the study of CBIR in medicine, has recently become an active area of research (Petrakis

* Corresponding author. Address: School of Information Technology, Building J12, University of Sydney, NSW 2006, Australia. Tel.: +61 2 9036 9800; fax: +61 2 9351 3838.
E-mail addresses: ashnil.kumar@sydney.edu.au (A. Kumar), jinman.kim@sydney.edu.au (J. Kim), lingfeng.wen@sydney.edu.au (L. Wen), michael.fulham@sydney.edu.au (M. Fulham), dagan.feng@sydney.edu.au (D. Feng).

1361-8415/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.media.2013.11.003

et al., 2002; Petrakis, 2002; Alajlan et al., 2008; Fischer et al., 2008; Huang et al., 2012; Xu et al., 2008; Lee et al., 2009; Iakovidis et al., 2009; Shyu et al., 1999; Aisen et al., 2003; Müller et al., 2005; Napel et al., 2010; Quellec et al., 2010a; Quellec et al., 2010b; Quellec et al., 2011; Akakin and Gurcan, 2012). In clinical evaluation, a lung image retrieval system (Shyu et al., 1999) assisted physicians to identify disease more accurately (Aisen et al., 2003). Müller et al. (2005) used a CBIR system to access images from a teaching database. Napel et al. (2010) concluded that CBIR could provide real-time clinical decision support to radiologists.

Features that describe relationships within images, such as the spatial arrangement of regions of interest (ROIs), are now being used by both general and medical CBIR studies. Hoàng et al. (2010) portrayed this using triangular spatial relationships, i.e., angles between groups of entities. Jaworska et al. (2010) showed the spatial arrangement using matrices that indicated the relative cardinal or ordinal (compass) directions of objects. The interactions between the contours and areas of two objects (such as overlap) were described using complex strings (Wang et al., 2012). In the medical domain, Huang et al. (2012) calculated the spatial relationships between adjacent objects through the intensity profile of the local neighbourhood of tumour ROIs. Spine X-ray retrieval has been achieved using partial shape matching (Xu et al., 2008), and the spatial and geometric features between adjacent vertebrae (Lee et al., 2009). Graphs are a more general approach to represent relational data (Bunke and Riesen, 2011). Retrieval methods based on the graph edit distance have high retrieval accuracy compared to other methods, such as strings (Petrakis, 2002). Alajlan et al. (2008) proposed a tree-based CBIR algorithm that used ROI curvature and hierarchical inclusion. Jouili and Tabbone (2012) reported a graph indexing technique that assigned graphs to multiple clusters. Fischer et al. (2008) used hierarchical attributed regional adjacency graphs to model the relationships between adjacent regions.

Despite advances in medical CBIR, there are still many challenges to overcome (Müller et al., 2004; Müller et al., 2007; Müller et al., 2009; Akgül et al., 2011; Ghosh et al., 2011; Müller and Deserno, 2011). One is the retrieval of multi-modality images, which are images of the same body region acquired with different techniques, such as computed tomography (CT) and positron emission tomography (PET), referred to as PET-CT. Multi-modality imaging introduces the possibility of searches based on complementary features from the different modalities. Recent studies on CBIR algorithms on different imaging modalities (Quellec et al., 2010a; Quellec et al., 2010b; Quellec et al., 2011; Akakin and Gurcan, 2012) still treated each image modality separately. Foncubierta-Rodríguez et al. (2013) quantified the texture information from CT images acquired at different energy levels; the combination of energy and spatial information was applied to the early diagnosis of pulmonary embolism. Zhou et al. (2012) reported a case-based retrieval algorithm that combined multi-image queries from different imaging modalities. Retrieval was achieved by finding the most similar images to every image within the query case and reducing the similar images to a list of unique cases in the dataset. However, the complementary information relationships between the different modalities within the cases were not exploited.

In this paper, we will explore CBIR as applied to multi-modality images acquired sequentially on a PET-CT scanner (Townsend et al., 2003). PET-CT provides complementary information from each modality and has an important clinical role in the evaluation of a variety of malignancies with improved diagnosis, staging, and assessment of treatment response when compared to the single imaging modalities (Blodgett et al., 2007). PET-CT can be regarded as the template for multi-modality imaging as other modalities, such as SPECT-CT and PET-MR, are introduced more widely into clinical care.

The clinical motivation for this work includes developing a methodology to retrieve image volumes from specific conditions for training and education and also for research. We have chosen a clinical dataset (Section 3) for patients with non-small cell lung cancer (NSCLC) as an example. Lymphoma could also have been chosen but we decided to limit our initial study to the thorax. In NSCLC, the most important determinant of outcome is the clinical stage at diagnosis as determined by the tumour location as well as the detection of regional nodal disease in the hilar and mediastinal regions (Detterbeck, 2009; Edge et al., 2010; Edge and Compton, 2010). The location of the tumour is also an important consideration in regard to the surgical approach to the tumour. Hence, the spatial arrangement of different tumours is critical in the accurate evaluation and staging of lung cancer. From a research perspective, the spatial mapping of different tumours provides important insights into the heterogenous manifestations of the disease and its patterns of spread. A CBIR system where image similarity is based on similar criteria could provide opportunities for comparing protocols, case-based reasoning, training, and comparison of data from clinical trials.

To the best of our knowledge, our preliminary investigations are the only work in the area of multi-modality PET-CT retrieval. We have previously investigated the feasibility of CBIR for PET-CT using different features (Kim et al., 2009; Song et al., 2010; Song et al., 2011). We have also represented the spatial relationships between functional (PET) and anatomical (CT) structures as complete graphs for both 2D slices (Kumar et al., 2008) and 3D volumes (Kumar et al., 2012b). We previously reported that the retrieval could be improved by reducing the graph edges (pruning) to remove redundant information (Kumar et al., 2012a). However, our previous graph-based method (Kumar et al., 2008; Kumar et al., 2012b) for PET-CT images used features common to both modalities and did not include modality-specific features such as CT texture or the standard uptake value (SUV), which in PET is a semi-quantitative measure of the most common PET tracer, ^{18}F-fluorodeoxyglucose (^{18}F-FDG).

However, representing all relationships on a complete graph obscures important relationships in multi-modality images. We hypothesise that in PET-CT studies of patients with cancer, we can enhance retrieval accuracy by constraining the graph representation of tumours to related anatomical structures and indexing complementary features that are unique to different modalities, i.e. using retrieval according to tumour localisation. This will mimic the geometric and topological attributes utilised in the staging system for cancer (Detterbeck, 2009; Edge et al., 2010), where disease classification is based upon the size, location, and the relationship of tumours to adjacent structures, e.g. the mediastinum in lung cancers, or the diaphragm, bone marrow, and spleen, in lymphomas (Carbone et al., 1971; Lister et al., 1989). As such, we propose a new representation for our CBIR framework, the Complete-Anatomy Proximal-Pathology (CAPP) graph, which is designed to prioritise spatial relationships between tumour and organ ROIs (extracted semi-automatically from two modalities). We also introduce a graph similarity measure that accommodates the modality-specific features indexed on the CAPP graphs. We applied our method to the retrieval of 3D PET-CT volumes.

2. Theoretical background and design

2.1. Graphs

Graphs are a general approach to representing symbolic structured data (Bunke and Riesen, 2011). A graph representation $G = (V, E, A, a, S)$ consists of a set of vertices V, which represent individual objects; a set of edges E that link vertices and represent

relationships between objects; an alphabet A, which is the set of attributes that describe the objects or relationships; a set of functions a for calculating and assigning the attributes to vertices or edges; and a symbolic reference S to the structure being represented, e.g. file path, object identifier. The *order* of a graph is the number of vertices.

Fig. 1(a) depicts a structure with four objects, each with a different colour. The graph in Fig. 1(b) represents each object as a vertex (circle) with the same colour as the object and a single attribute (α) describing its area. Relationships between objects are represented by edges (lines) with a single attribute describing the distance between the objects (δ). The set of functions a was used to calculate the vertex and edge features (α and δ). Finally, the graph contains a file path (S) that links it to the structure from which it was derived, shown in Fig. 1(a).

In such graph representations, each individual vertex or edge can be treated as an independent feature vector, i.e., a vector $\mathbf{v} \in V = [x_1\ x_2\ \ldots\ x_k]$ and an edge $\mathbf{e} \in E = [y_1\ y_2\ \ldots\ y_l]$, for feature values given by x_i where $1 \leqslant i \leqslant k$ and y_j where $1 \leqslant j \leqslant l$. The graph is the collection of these independent feature vectors.

Graphs have represented relational and structural information in many pattern recognition applications, including image matching (for a review see Conte et al., 2004). These representations demonstrated high retrieval accuracy when searching for images based upon the similarity of the spatial arrangement of their ROIs (Petrakis, 2002). Furthermore, graphs are able to index any possible feature as an attribute, and unlike feature vectors are not constrained to a fixed length (Bunke and Riesen, 2011).

Graphs can be compared in terms of structure and the attributes indexed on the vertices and edges. The graph edit distance is a measure of the dissimilarity of two graphs (Bunke, 1999) that measures the cost to transform one graph (G) into another graph (H). Let $o = \langle g, h \rangle$ be an edit operation, where g is a vertex (or edge) of G and h is a vertex (or edge) of H. When g and h are not null ($\varnothing$) then o is a substitution operation; when g or h is null then o is insertion or deletion. Using this terminology, the graph edit distance can then be defined as:

$$D_{GED}(G,H) = \min_{(o_1,o_2,\ldots,o_n)\in\Omega} \sum_{i}^{n} d(o_i) \tag{1}$$

where $(o_1, o_2, \ldots, o_n)$ is a sequence of edit operations, Ω is the set of all sequences that transform G to H, and $d(o_i)$ is a function for calculating the cost for the edit operation o_i.

The major drawback of graph representations comes from the high computational complexity of many graph algorithms, including the most accurate 'brute force' methods for calculating the graph edit distance (Petrakis, 2002; Bunke and Riesen, 2011), which we used previously (Kumar et al., 2008). The brute force method finds the optimal sequence of edit operations by evaluating every element (unique sequence of edit operations) in Ω. A tree structure is iteratively generated to enumerate the unique sequences; each node is a pair of mapped vertices with each path from the root to a leaf node representing a unique sequence. As the order of the graphs grow, so does the length and number of sequences, and the time needed to evaluate the graph edit distance.

While faster methods are generally less accurate, Neuhaus et al. (2006) showed that their 'beam search' approach was almost as accurate as a brute force method when using large beams. The beam search algorithm is a fast but suboptimal algorithm for calculating the graph edit distance through an adaptation of the popular A* algorithm (Hart et al., 1968). Unlike the brute force algorithm, the beam search does not evaluate every sequence in Ω. Instead, it generates a smaller set of sequences $\Omega_b \subseteq \Omega$ by iteratively expanding the b best partial branches (sequences) in the tree. This speeds up the operation by limiting the number of sequences that have to be evaluated. The beam search algorithm was appropriate for ranking images based on similarity of their classes because it increased inter-class differences while intra-class differences were not strongly affected (Neuhaus et al., 2006).

2.2. Proposed graph-based CBIR method

The major components of our graph-based CBIR framework are shown in Fig. 2. The pink arrows show the preprocessing used to create the search index. The images in the database are segmented, features are extracted from the segments, and represented on graphs; the graphs are stored in the index. The green arrows show the query process. First, a graph is created from the query image (after segmentation and feature extraction). The query graph is then compared to the indexed graphs. Finally, the most similar images, according to the graph comparisons, are displayed to the user.

In most graph-based CBIR frameworks (Petrakis et al., 2002; Petrakis, 2002; Alajlan et al., 2008; Kumar et al., 2008; Kumar et al., 2012a), ROIs within the images are represented by graph vertices. Vertex attributes are used to index image features, for example ROI curvature (Alajlan et al., 2008), or sets of ROI features, including size, orientation, or perimeter (Petrakis, 2002). Relationships between two ROIs are represented by the graph edge that connects their corresponding vertices. These edges can represent binary relationships, such as inclusion (Alajlan et al., 2008), or sets of attributed relationships (distance and angle between the two ROIs) (Petrakis, 2002). The similarity of images is calculated by finding the graph edit distance where the cost of an edit operation ($d(\cdot)$ in Eq. 1) is based on the differences in graph structure and attributes.

In this work, an image is considered relevant to the query if the two images share the same tumour localisation, i.e., the tumour location relative to organs is the same. The set of geometric and topological attributes that represent the anatomical location extent of tumours can be formulated as a graph. Our choice of graphs was motivated by their ability to represent the spatial arrangement of multiple ROIs (tumours and organs) as well as other relationship

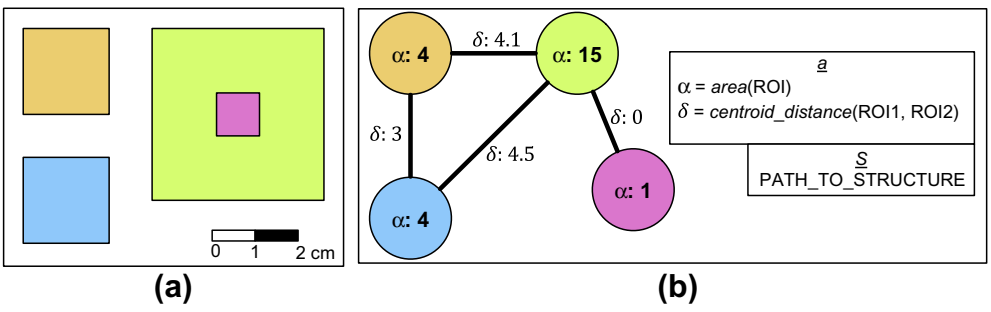

Fig. 1. Representing a structure using graphs. (a) A structure with four objects. (b) The associated graph representation with attributes describing the area of each object (α), and the distance between objects (δ); these attributes were calculated using the set of functions given by a. The graph also contains a link (S) to the original structure.

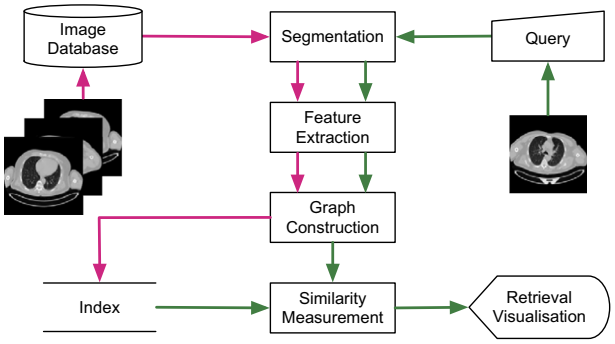

Fig. 2. Our graph-based CBIR framework. The pink arrows depict the creation of the image index and the green arrows show the query process. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

features in multi-modality images. An example of the clinical relevance of spatial arrangement of tumours relative to anatomy is in the staging of primary lung tumours (Detterbeck, 2009): Stage T1 is assigned to cases with a tumour that is completely enclosed by the lung while Stage T3 is assigned to cases with a tumour that is invading the chest wall, diaphragm, or mediastinal pleura. Tumour properties (size etc.) are used to further stage the disease into different classes. While spatial arrangement can be modelled by triangular relationships (Hoàng et al., 2010) or matrices describing the relative placement of objects (Jaworska et al., 2010), these methods cannot present other relational information, such as the relative size of two objects or the overlap between them. Strings can describe the interactions between ROIs (Wang et al., 2012) but Petrakis (2002) has shown that graph based methods are more accurate for retrieval based on the relative spatial arrangement of ROIs.

Our choice of features (see Section 4.3) was based on the need to represent the geometric and topological attributes used for cancer staging and classification (Detterbeck, 2009; Edge et al., 2010; Carbone et al., 1971; Lister et al., 1989), including geometric tumour properties (volume, length, etc.) and location relative to anatomy (distances and angles from other structures). To this end, we adapted the geometric and spatial features described by Petrakis (2002) and complemented them with modality specific information, SUV for PET and texture for CT. This method differs from our previous work (Kumar et al., 2008; Kumar et al., 2012b) where features common to both modalities were used.

Our CBIR framework is predicated on the following assumptions:

1. The two modalities present both common and complementary information; anatomical data are provided by CT, and functional or tumour data from PET.
2. The scanner hardware co-aligns the images. In this study, we disregard errors in aligning regions between modalities (registration).
3. The segmentation algorithms do not label the extracted ROIs. As such, we assume that the extracted structures are unknown. Only the modality of the image from which ROIs are extracted is known.
4. Each ROI was a collection of voxels belonging to a tumour or an anatomical structure.

3. Materials

3.1. Clinical PET-CT dataset

We used 50 PET-CT studies of lung cancer patients that were acquired on a Siemens Biograph mCT scanner, each comprised of two series: one CT volume and one PET volume. The reconstructed volumes had a CT resolution of 512×512 pixels at $0.98\,\text{mm} \times 0.98\,\text{mm}$, a PET resolution of 200×200 pixels at $4.07\,\text{mm} \times 4.07\,\text{mm}$, and a slice thickness and an interslice distance of 3 mm. The studies in the dataset contained between 1 and 7 tumours (inclusive). The clinical reports detailing tumour locations and nodal involvement were available for each study.

We used a well-established adaptive thresholding algorithm with refinements (Hu et al., 2001) to segment the left and right lung volumes from the CT volume. Tumours from the PET volumes were segmented with a 40% peak SUV connected thresholding to detect 'hot spots' identified in the diagnostic reports (Bradley et al., 2004). To include other major organs above the diaphragm we applied connected thresholding to coarsely segment the brain and mediastinal tissue (including the heart).

We used the findings from the clinical reports to make minor manual adjustments to ensure that the segmented ROIs were accurate. The manual interventions were generally required for over-segmentation (extra or larger ROIs), rather than under-segmentation (missed or smaller ROIs). The adjustments included:

- eliminating PET ROIs that were not tumours; extraneous ROIs were selected for deletion prior to graph creation and this was achieved with 2 mouse clicks per extra ROI;
- adjusting the parameters for coarse segmentation of the mediastinum and brain by manual entry of upper and lower threshold values or dragging two sliders; and,
- separating the left and right lungs in 2 of the 50 studies; for these it was only necessary to split the segmented lung ROIs in a few slices because they were only joined at the anterior part of the thorax.

3.2. Ground truth

The ground truth for our dataset was derived from the accompanying clinical reports. These reports were written by an experienced specialist physician related to the project who has read over 60,000 PET-CT studies. The tumour locations and nodal involvement stated in the clinical reports were used as labels for the associated images. A computer scientist related to the project collated these labels to construct the ground truth. That is, the tumour locations and nodal involvement were used as expertly-assigned labels and our retrieval method was to be evaluated against these labels.

4. Multi-modality graphs

Our method begins by extracting ROIs from each modality separately as described in Section 3. During graph construction, a vertex is created for each ROI and labelled with the image modality (anatomical or pathological) of the represented ROI. The attributed feature set of each vertex depends upon this label. Our attributed vertices index ROI features while our edges index relationships between ROIs. Edges are created between each vertex to form a complete graph (see Petrakis et al., 2002, for an example). We created CAPP graphs by pruning edges from the complete graphs based upon the physical spatial proximity of tumours to organs. The extensibility of graph representations (Section 2.1) enabled us to index modality-specific features. Furthermore, when images contained multiple tumours and organs, additional vertices and edges were included to represent each structure individually.

4.1. Complete graphs

Let V_A be the set of all vertices representing anatomical ROIs. Similarly, let V_P be the set of all vertices representing tumour ROIs. Let F_A, F_P and F_S be the feature alphabets for anatomical regions,

pathology (tumour) regions, and spatial relationships, respectively. Similarly, let f_A, f_P and f_S be the functions that calculate these features and assign them to elements of V_A, elements of V_P, and graph edges, respectively.

We can then define the multi-modality complete graph $G_C = (V_C, E_C, F, f, I)$, where $V_C = V_A \cup V_P$, $E_C = \{v_i v_j\} \forall v_i,\ v_j \in V_C$ where $v_i \neq v_j$ is the set of all edges, $F = (F_A, F_P, F_S)$ is the combined feature alphabet, $f = (f_A, f_P, f_S)$ are the functions for calculating these features, and I is the multi-modality image represented by G_C. By definition, all complete graph vertices are pairwise connected by an edge.

The construction of a complete graph representing a PET-CT image is shown in Fig. 3. Features are extracted from the ROIs in each modality and indexed on vertices. The vertices have been labelled with the modality of the image from which the ROIs are extracted. Alignment of the two images by the scanner hardware allows the extraction of spatial relationships between ROIs in different modalities, thereby enabling the construction of the complete graph.

4.2. Complete-anatomy proximal-pathology graphs

The Complete-Anatomy Proximal-Pathology (CAPP) graph is the representation we propose to constrain the pathology (tumours) to spatially related anatomical structures. Let $G_{CAPP} = (V_C, E_{CAPP}, F, f, I)$ be the CAPP graph where $E_{CAPP} \subseteq E_C$. We generate E_{CAPP} by pruning the complete edge set E_C using an indicator function Φ. Let $v_i, v_j \in V_C$ with $v_i \neq v_j$ be the endvertices of an edge $v_i v_j \in E_C$. The function Φ uses the modalities and proximity of v_i and v_j to signal inclusion within E_{CAPP}. We define Φ as follows:

$$\Phi(v_i v_j) = \begin{cases} 1 & \text{if } v_i \in V_A \text{ and } v_j \in V_A \\ & \quad v_i \in V_A \text{ and } v_j \in V_P \text{ and} \\ 1 & \text{if } v_i = \underset{x \in V_A}{\arg\min}\ proximity(x v_j) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $proximity(xy)$ is a function that obtains the spatial nearness of the regions represented by vertices $x \in V_C$ and $y \in V_C$ from the edge $xy \in E_C$. We define $proximity(xy)$ as returning the md feature of edge xy (see Section 4.3 for an explanation of the different features). This definition enables CAPP graphs to model tumour localisation.

Fig. 4 depicts the creation of the CAPP graph of the PET-CT image in Fig. 3. The function Φ indicates the edges that are to be preserved: the green edge connects two vertices that are elements of V_A, while the purple edges connect vertices in V_P to the vertex in

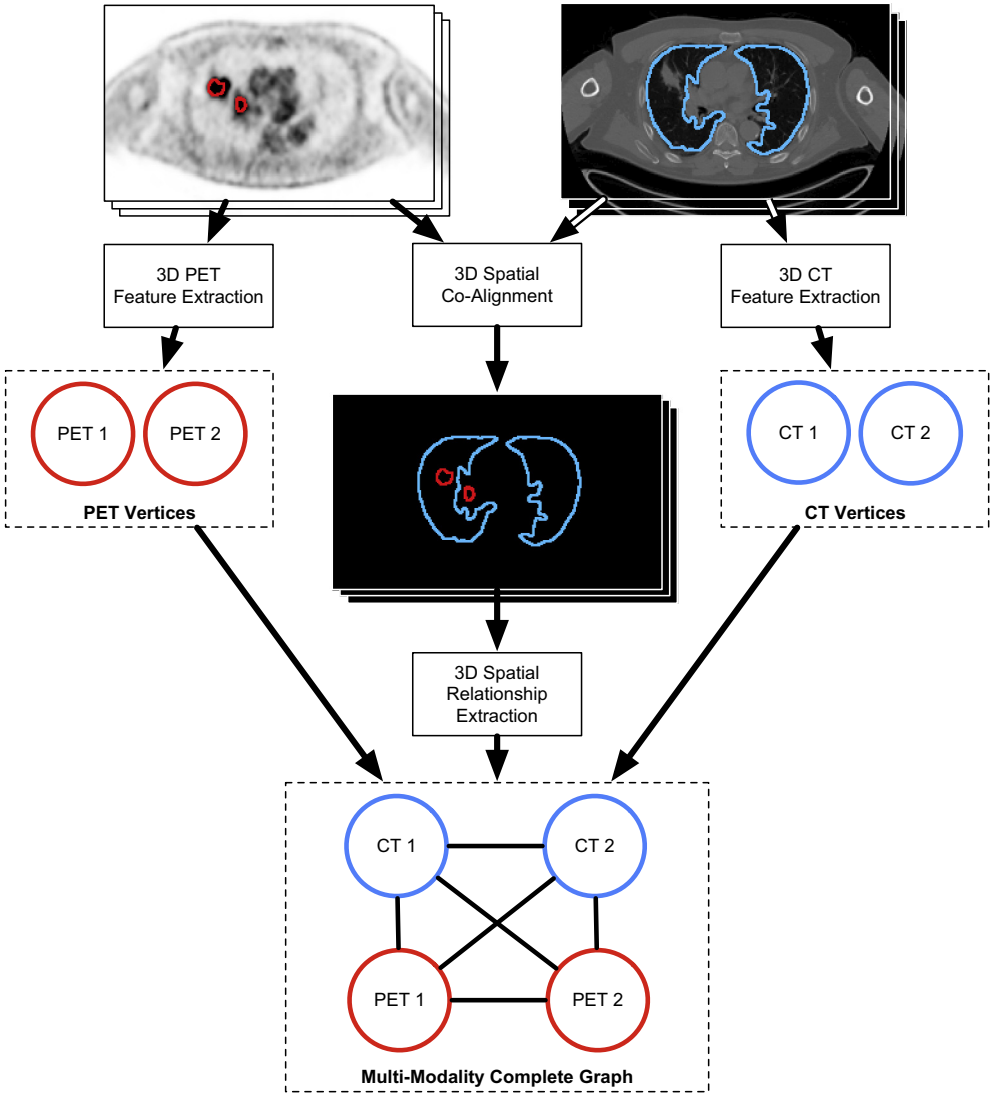

Fig. 3. Multi-modality complete graph construction. Features from ROIs in both modalities are indexed on graph vertices. Alignment of the two modalities by the scanner hardware is used to calculate spatial relationships between all ROIs (including those in different modalities). These relationships form the graph's edges.

V_A representing the spatially nearest anatomical ROI. The edges indicated by dashed lines have been removed in the resultant CAPP graph.

Our definition of the edge set of the CAPP graph is similar to edge set of the regional adjacency graph (RAG) (Llados et al., 2001). The edges in an RAG are incident to vertices representing ROIs that are adjacent within an image. However, because the ROIs in different modalities may not be directly adjacent to each other (due to segmentation, partial volume effects etc.), proximity is used for CAPP graph edge definition.

4.3. Image features

We divided the feature set into three categories, corresponding to the two vertex sets and the edge set: anatomical ROI features (F_A), tumour ROI features (F_P), and spatial relationship features for edges (F_S). That is, we treated each vertex and edge on a graph as individual feature vectors: vertices in V_A were $|F_A|$-dimensional vectors, vertices in V_P were $vertF_P$ |-dimensional vectors, and edges were $|F_S|$-dimensional vectors. In this section, we list all the features extracted for our experiments.

We indexed the following features as vertex attributes: size (s), boundary (b), length or the maximum distance between two points on the boundary of a ROI (l), roundness (r), and tumour homogeneity (th). The maximum (I_{MAX}), mean (I_μ), and variation (I_σ) of the intensity or SUV of pixels in an ROI were also indexed as vertex features. The graph edge attributes were: distance (d), relative orientation (ro), relative size (rs), and minimum distance (md).

Size was represented by the volume (number of voxels) of a particular ROI. Similarly, the boundary size was defined as the surface area of a 3D ROIs. The ro feature consisted of two values, equivalent to the pitch and yaw angles directing an observer from the centroid of one ROI towards the centroid of the other. The roundness or sphericity of a 3D ROI was defined as a function of its volume (the s feature) and surface area (the b feature), as in Wadell (1935).

Tumour homogeneity is a measure of the uniformity of tumour pixel intensities. We adapted the voxel neighbourhood homogeneity coefficient given by Belhassen and Zaidi (2010), which considers intensity distribution with spatial constraints. Our adaptation limited the calculation of the coefficient to only neighbouring voxels within the tumour ROI instead of all 26 3D neighbours. Let $P = \{p_1, p_2, \ldots, p_n\}$ be the set of voxels within a tumour. The tumour homogeneity is then given by:

$$th(P) = \frac{\sum_i \alpha(p_i)}{n} \tag{3}$$

and

$$\alpha(p) = \frac{1}{1 + \sqrt{\frac{1}{|nd(p)|}\sum_{k=1}^{|nd(p)|}(p_k - p)^2}} \tag{4}$$

where $p_i \in P, |\cdot|$ is the cardinality function, and the function $nd(p)$ returns a set of voxels $P_{nd} \subset P$ that are the neighbours of p, and $p_k \in P_{nd}$ is a voxel that is a neighbour of p.

Thirteen gray-level co-occurrence matrices (one for each unique direction) were used to calculate 3D Haralick texture features. The matrices were derived from the co-occurrence of neighbouring voxel values. We calculated five features from these co-occurrence matrices (Haralick et al., 1973): entropy (ent), contrast ($cont$), correlation (cor), energy (nrg) and homogeneity ($hmgt$).

Acquisition protocols that use the same patient orientation (e.g. head-first and supine) for images with the same resolution enable the extraction of point sets (pts), the coordinates of all the pixels in a given ROI. We used this to improve anatomical matching by measuring the overlap between anatomy ROIs in the query and dataset images, as in Kim et al. (2009).

4.3.1. Feature selection

In this work, we did not find the optimal set of features using feature selection. This is because feature selection is regarded as a separate, significant challenge in graph-based CBIR studies, which are instead focused on presenting new ways for graph construction and matching (Petrakis, 2002; Alajlan et al., 2008; Jouili and Tabbone, 2012). The graphs in these studies, as well as in ours, represented both structural and image features. Any feature selection technique would therefore need to optimise the structures, the features, and their interdependence. For example, in our context the optimality of a feature may be dependent on a number of factors: the modality the feature was extracted from, the structure on which it is indexed, and other related structures in the same graph. To the best of our knowledge, there is no standard method for performing this type of feature selection.

We therefore chose to use a standard set of spatial and relational features (Petrakis, 2002) that were complemented with modality specific information. We also normalised the features (Section 5.1) to ensure that the contribution of each feature was independent of its range of values.

5. Calculating graph similarity

5.1. Feature normalisation

Our graphs indexed various types of features and each feature had a range of different of possible values. As such, there was a risk that features with a large range would dominate the similarity matching compared to features with small ranges. Aksoy and Haralick (2001) demonstrated that linearly scaling features to a random variable with zero mean and unit variance improved the discriminatory capabilities of similarity measures. However, this approach could not be applied directly to all the features indexed on our graphs. As such, we derived a different normalisation technique for different features depending on the type: measurements

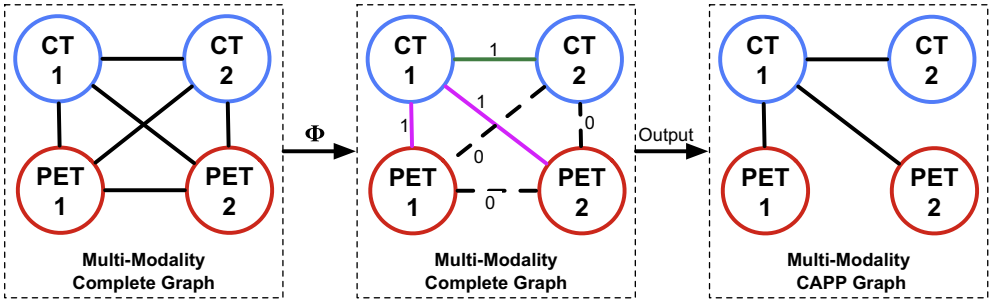

Fig. 4. The CAPP graph of the PET-CT image from Fig. 3; Φ retained the green edge as it was incident to two vertices in V_A; the purple edges are retained because they connected pathology vertices to the nearest anatomical vertex. Dashed edges were eliminated in the final CAPP graph. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(such as volume, length, etc.), angles, and point sets. Each of the normalisation techniques ensured that the contribution of each feature to the similarity measure was independent of its range of values and was within the range $[0,1]$.

5.1.1. Measurement features

Measurements are features that represent a certain property such as distance or size. Thus, it is possible to calculate the mean and standard deviation of all the feature values in the dataset and scale the features accordingly. However, in our multi-modality cases it was possible for the range of features to be different across modalities (e.g., tumour volume can be much smaller than organ volume). Hence, this normalisation should be applied separately for different modalities, e.g., anatomical volume should be scaled using the mean and standard deviation of the volume of anatomical ROIs.

Let x be the value of a feature f, and μ_f and σ_f be the mean and standard deviation of f in the dataset. We normalised different values of x to the range $[0,1]$ by linearly scaling them to a random variable with zero mean and unit variance and shifting it so it is within the desired range. The normalised value $\tilde{x}$ of x was determined by using a function from Aksoy and Haralick (2001):

$$\tilde{x} = \frac{(x-\mu_f)/3\sigma_f + 1}{2} \tag{5}$$

5.1.2. Angular features

While angular features (such as relative orientation) are also measurements, the circular nature of the angles makes them difficult to normalise by linear scaling as in Section 5.1.1. The difficulties arise when normalising angles that have a large difference but are similar when plotted on a circle. For example, the values $+(\pi-\epsilon)$ and $-(\pi-\epsilon)$, for a small value $\epsilon > 0$ have a difference of $2(\pi-\epsilon)$ but the angles are only 2ϵ radians apart.

We therefore normalised an angular feature θ using functions of its *sine* and *cosine* (Petrakis, 2002). The normalised value $\tilde{\theta}$ was a pair of values:

$$\tilde{\theta} = \left\langle \frac{\sin(\theta)+1}{4}, \frac{\cos(\theta)+1}{4} \right\rangle \tag{6}$$

The maximum value of each component was 0.5. This ensured that the contribution of a single angle lay within the range $[0,1]$.

5.1.3. Point set features

Point set features were collections of coordinates representing the shape and layout of the ROI, such as the contour or boundary points. We normalised the difference between two point sets during similarity measurement such that the distance value ranged from 0 (total similarity) to 1 (total dissimilarity). The Jaccard distance was used to measure the dissimilarity between two point sets:

$$distance(q_{pts}, s_{pts}) = 1 - \frac{|q_{pts} \cap s_{pts}|}{|q_{pts} \cup s_{pts}|} \tag{7}$$

where q_{pts} and s_{pts} are two point sets of a query and dataset vertex, respectively. The distance value was within the range $[0,1]$.

5.2. Graph similarity

We calculated similarity measures by calculating the graph edit distance using an adaptation of the beam search algorithm (Neuhaus et al., 2006). Both the standard brute force approach and its beam search adaptation assume a standard set of vertex features. The complete and CAPP graphs have different feature sets for different graph vertices and edges (F_A for V_A, F_P for V_P, and F_S for edges). It was necessary to adapt the beam search algorithm to account for these cases. Since organs are not tumours, we achieved this by disallowing comparisons between different modalities using a specially designed cost function for our graph edit distance calculation (Eq. 1):

$$d(\langle Q,S\rangle) = \begin{cases} \infty & \text{if } mod(Q) \neq mod(S) \\ \left[\sum_i (q_i)^p\right]^{\frac{1}{p}} & \text{if } S = \varnothing \\ \left[\sum_i (s_i)^p\right]^{\frac{1}{p}} & \text{if } Q = \varnothing \\ \left[\sum_i (q_i - s_i)^p\right]^{\frac{1}{p}} & \text{otherwise} \end{cases} \tag{8}$$

where $\langle Q,S\rangle$ was a graph edit operation, Q was a vertex (or edge) from the query graph, S was a vertex (or edge) from a graph in the dataset, *mod* was a function that returns the modality of a vertex or edge, q_i was the ith feature of Q, s_i was the ith feature of S, and p was the order of the distance function. When the ith feature was a point set, the difference $q_i - s_i$ was replaced with the result of Eq. 7. In our experiments, we set $p = 2$. This was equivalent to the Euclidean distance except in the case of point sets.

The first case in Eq. 8 assigns an infinite cost when the transformation sequence attempts to substitute vertices that are from different modalities (i.e., transforms an organ into a tumour or vice versa). The infinite cost ensures that the minimisation part of the graph edit distance calculation (see Section 2.1) eliminates the sequence from consideration because of this single operation. This forces the algorithm to consider other alternate sequences in Ω, possibly involving the deletion or insertion of Q or S.

The second and third case cover the cost of deletion or insertion operations, respectively (Petrakis et al., 2002). The second case measures the cost of deleting Q from the query graph while the third case measures the cost of inserting S into the query graph. An alternate but equivalent formulation is that the second case measures the cost of inserting Q into the graph from the dataset while the third case measures the cost of deleting S from the graph in the dataset. The final case is the cost of substitution.

The differences between Eq. 2 and Eq. 8 are important to note. Eq. 2 applies to the construction of CAPP graphs and it allows edges between vertices (ROIs) from different modalities to enable relationships between tumours and organs to be represented on a *single* graph. On the other hand, Eq. 8 prevents substitution of vertices from *different* graphs if the vertices are from different modalities.

6. Experimental procedure

Our experiments were designed to determine if our method enabled the retrieval of multi-modality images by comparing the retrieval results to the ground truth. We evaluated our work using two standard measures: precision, the proportion of retrieved images that were relevant; and recall, the proportion of all relevant images in the dataset that were actually retrieved. We also calculated the mean average precision (MAP), a single-value indicating the average precision over all levels of recall. The relevance of a retrieved image was determined using the ground truth described in Section 3.

Our retrieval experiments used a leave-one-out (LOO) cross-validation approach. We divided the dataset of k images into k sets consisting of 1 query image and $k-1$ indexed images. All optimisations, such as calculating the mean and standard deviation for feature normalisation or extracting SIFT visual words (see below), were performed on the set of $k-1$ indexed images. The removal

of the query from the indexed images ensured that our optimisations as well as the final precision and recall were not biased by the presence of the query. We compared our CAPP graph retrieval method to the following techniques:

- Gray-level histograms extracted from the thorax of PET-CT 3D volumes. One 256-bin histogram was computed for each of the modalities. This is a traditional image retrieval technique (Bugatti et al., 2009).
- Visual words using SIFT features (see Zhou et al., 2012 for a description). The features were extracted from the PET and CT slices showing the thorax on a slice-by-slice basis; the features were combined to form descriptors for the 3D volume. This is a state-of-the-art retrieval method that does not use spatial relationships.
- Relational matrices representing spatial relationships between ROIs. This is similar to the technique suggested by Jaworska et al. (2010) except it was adapted for 3D volumes. This is an advanced retrieval method that is based on spatial relationships.

We also compared our CAPP graph to the complete graph to investigate whether constraining tumours to their nearest anatomical structures improved the precision of the image retrieval algorithm. We also investigated the robustness of our graph-based method when fully automatic segmentation was used.

Finally, we examined the contribution of the image features on our graph-based PET-CT retrieval. We first created multiple new feature sets, each of which excluded a feature from the full feature set F. That is, we constructed new feature sets $F_i = F \setminus f_i$ where $f_i \in F$ and $1 \leqslant f \leqslant |F|$. The MAP of the LOO experiments using each F_i was then calculated. We also investigated the contribution of the graph structure by repeating the LOO experiments using no image features.

7. Results

Fig. 5 shows a comparison of the mean precision and recall achieved by our graph-based CBIR method and the other methods. Table 1 summarises the MAP of these results and shows that the CAPP graph had a higher overall mean precision than all other methods. We also examined the performance of our method in retrieving images with tumours in particular locations. These results are given in Table 2; the bold values indicate the highest MAP across the different methods for a specific location. The CAPP graph had higher a MAP when retrieving tumours in the majority of anatomical locations. The MAP values in Table 2 are generally lower than those in Table 1 because queries with multiple tumours often resulted in the retrieval of images with tumours that shared other localisations with the query but not necessarily the one that was targeted.

We evaluated the robustness of our method to segmentation errors by experiments using representations where the ROIs were detected by automatic segmentation. Fig. 6 depicts the mean precision and recall of the experiments when using automatic segmentation techniques. The MAP of these experiments are summarised in Table 3, which shows that the CAPP graph achieved a higher overall MAP than the other methods. While the complete graph achieved a higher average precision at recall $< 10\%$, it had a lower precision for recall $\geqslant 10\%$. However, the higher MAP of the CAPP graph demonstrates that it has a better overall performance than the complete graph. Table 4 provides a breakdown of the MAP by tumour location; the bold values indicate the highest MAP across the different methods for a specific location. The MAPs of the SIFT and histogram methods, which are not reliant upon segmentation, are the same as in Table 2. The CAPP graph achieves a higher MAP for more tumour locations than any of the other techniques.

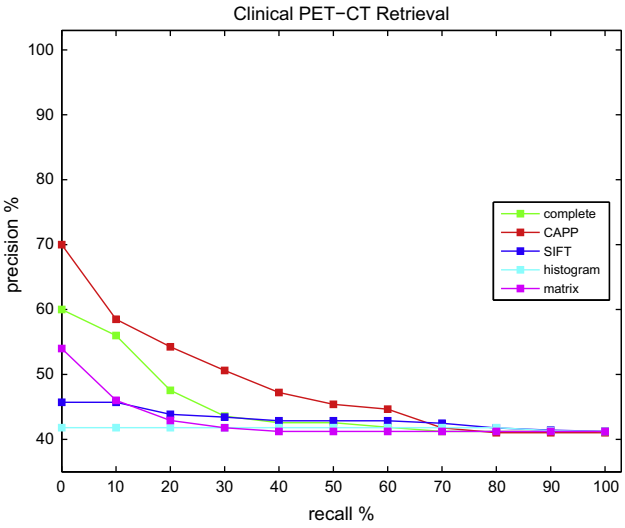

Fig. 5. Precision and recall graph averaged over all 50 queries.

Table 1
Comparison of mean average precision (MAP).

Method	MAP (%)
Complete	48.38
CAPP	52.28
Histogram	43.63
SIFT	46.52
Matrix	46.96

Table 5 lists the effect of the image features on the retrieval process; we calculated the MAP of CAPP graph searches using all features, using no features, and using our varied feature set (excluding one feature at a time). The table shows that the CAPP graph was able to maintain similar levels of retrieval precision using different feature sets.

A typical PET-CT retrieval example is depicted in Fig. 7. Table 6 gives the the locations of tumours in query and the top three retrieved results (R1, R2, R3) using the different methods. The query image contains a tumour within the lower lobe of the right lung with nodal involvement at the right hilum on the right side of the mediastinum (paratracheal lymph node involvement). All the images retrieved by the CAPP graph technique contained tumours localised in the right lower lobe, right hilum, or right mediastinum. Only R1 and R3 with the complete graph method retrieved images where the disease was in the right lung or right hilum. Only R2 with the gray-level histogram method showed a

Table 2
MAP by tumour location.

Location	MAP (%)				
	Complete	CAPP	Histogram	SIFT	Matrix
Right upper lobe (RUL)	50.70	50.27	41.64	45.54	**55.84**
Right middle lobe (RML)	12.51	12.20	14.44	**16.53**	10.85
Right lower lobe (RLL)	**24.51**	24.18	17.12	19.02	21.94
Left upper lobe (LUL)	14.45	**23.17**	20.05	19.64	11.45
Left lower lobe (LLL)	25.06	26.78	20.47	**31.78**	31.19
Mediastinum (M)	36.64	**39.14**	39.09	37.97	35.67
Right mediastinum (RM)	14.00	**24.68**	18.29	12.58	9.35
Right hilum (RH)	37.32	**45.52**	31.14	30.78	26.33
Left hilum (LH)	30.14	**45.30**	25.51	26.26	20.56

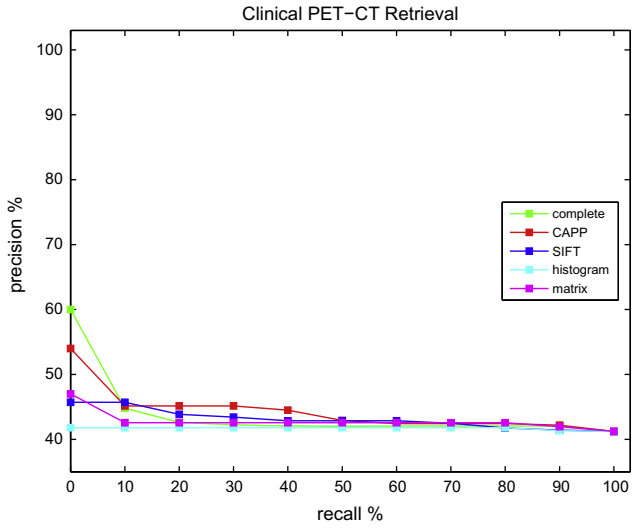

Fig. 6. Precision and recall graph averaged over all 50 queries (automatic segmentation).

Table 3
Comparison of MAP (automatic segmentation).

Method	MAP (%)
Complete	46.47
CAPP	47.53
Histogram	43.63
SIFT	46.52
Matrix	45.79

Table 4
MAP By tumour location (automatic segmentation).

Location	MAP (%)				
	Complete	CAPP	Histogram	SIFT	Matrix
RUL	42.44	42.95	41.64	45.54	**47.27**
RML	**19.30**	13.47	14.44	16.53	16.45
RLL	20.10	**22.94**	17.12	19.02	22.52
LUL	14.39	20.26	20.05	19.64	**22.72**
LLL	23.14	22.85	20.47	**31.78**	18.79
M	**42.85**	42.16	39.09	37.97	37.09
RM	21.33	**30.17**	18.29	12.58	13.23
RH	37.20	**48.25**	31.14	30.78	33.16
LH	28.57	**31.54**	25.51	26.26	25.92

lesion on the right that involved nodal disease. The SIFT and matrix methods also retrieved a single similar image and both methods retrieved misleading and inaccurate images. Our CAPP graph method retrieved PET-CT volumes that were realistic and most consistent with the patterns of disease found in clinical practice.

8. Discussion

Our findings show that our CAPP graph method had higher precision than the other methods when retrieving clinical PET-CT volumes of lung cancer patients. In the majority of the cases, the CAPP graph had a higher precision in retrieving image volumes with particular tumour localisation and the retrieved images were more similar to the query than those retrieved by the other methods.

The MAP results in Table 1 revealed an interesting pattern. The methods that used spatial relationships (complete graphs, CAPP graphs, and the matrix method) achieved higher MAPs than the techniques that did not use relationships. The graph-based

Table 5
MAP with feature removal.

Vertex feature	MAP (%)
Size (s)	52.44
Boundary (b)	52.50
Length (l)	52.47
Entropy (ent)	52.34
Contrast ($cont$)	52.35
Correlation (cor)	52.24
Energy (nrg)	52.27
Homogeneity ($hmgt$)	52.29
Roundness (b)	52.54
Tumour homogeneity (th)	52.35
SUV maximum (I_{MAX})	52.52
SUV mean (I_{μ})	52.42
SUV variation (I_{σ})	52.51
Point set (pts)	50.20
Edge feature	
Distance (d)	52.68
Relative size (rs)	53.40
Relative orientation (f)	48.40
Minimum distance (md)	52.20
All features	
CAPP – all used	52.28
Complete – all used	48.38
CAPP – none used	46.12

methods (complete and CAPP) were more precise than the relational matrix. The explanation is that the matrices only represent the relative position of two ROIs as one discrete descriptor from a predetermined set (e.g., above/below, left/right, front/behind, or a combination). Graphs, however, can represent relative position using the distance and the angle between objects as well as other relational information, such as the relative size of two ROI. Thus, graphs have a higher discriminatory power than the relational matrix.

Table 1 also shows that the CAPP graph has a higher MAP than the complete graph. This can be attributed to the variation in structure of CAPP graphs, which emphasised tumour localisation in multi-modality images by strongly associating individual pathologies and their nearest organs. Complete graphs had no variation in structure among graphs with the same number of vertices and as such they had less discriminatory power than CAPP graphs. This is illustrated in Fig. 8. The first column (a) depicts several images with grey anatomical regions and white tumours. Each of these images has the same number of ROIs. The same complete graph structure represents all these images (column (b)). The CAPP graphs in the third column (c) have different structures.

The variation in structure was introduced by Eq. 2, which created CAPP graphs by preserving edges between tumour vertices and the vertex of its nearest anatomical neighbour. Modifying Eq. 2 to include more edges would cause a decrease in retrieval precision, e.g., preserving edges with the two nearest anatomical neighbours reduces the MAP to 50.51%. The explanation for this effect is that increasing the number of edges (to the n closest organs) moves the CAPP graph's structure closer to that of a complete graph, and thus reduces its discriminatory power.

The discriminatory power of the CAPP graph also relates to the number of vertices (order). A higher order means that more varied graph structures are possible (see Table 2). The CAPP graph has a higher MAP than the other methods for tumours in the majority of anatomical locations. In each of these cases (LUL, M, RM, RH, and LH), more than 85% of the images in the dataset had multiple tumours. The MAP of the relational matrix method fluctuated depending on the tumour location. It was relatively high for the cases where the tumours were in locations that were in the 'corners' or the middle of the thorax, e.g. right upper lobe (top right),

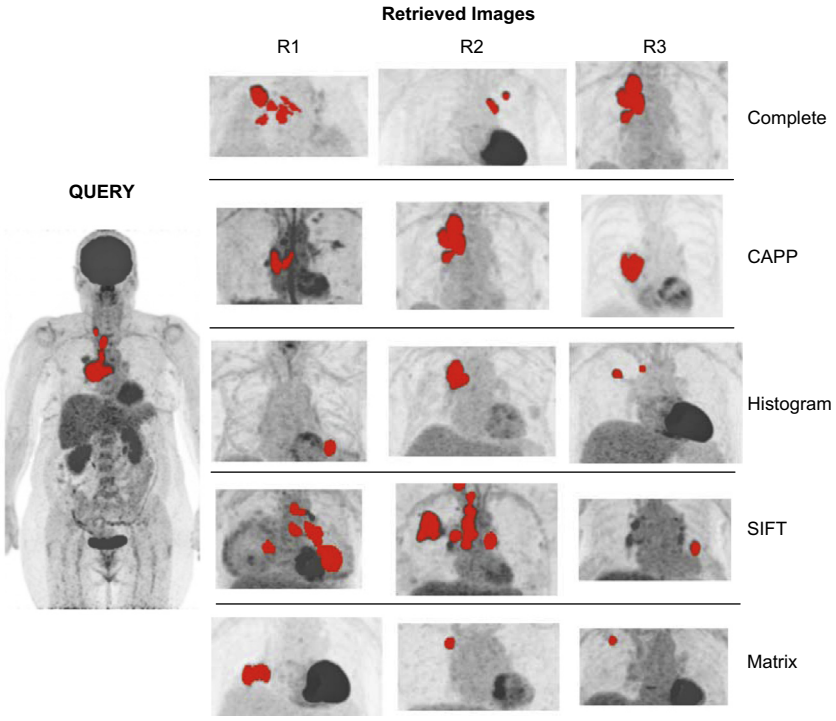

Fig. 7. The top 3 (R1, R2, R3) PET-CT volumes retrieved by the different methods. The figure shows maximum intensity projections (2D projections of the highest intensity 3D voxels) of the PET images; tumour ROIs are marked in red; retrieved images have been cropped to the lung fields. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
Tumour localisation for Fig. 7.

	Localisation labels[a]			
	1	2	3	4
Query	RLL	RH	–	–
Complete				
R1[b]	RUL	RH	M	–
R2	LUL	LH	M	–
R3	RUL	RH	M	–
CAPP				
R1	M	RH	–	–
R2	RUL	RH	M	–
R3	RLL	RH	–	–
Histogram				
R1	LLL	–	–	–
R2	RUL	M	–	–
R3	RUL	–	–	–
SIFT				
R1	LLL	LH	LM	–
R2	RUL	RH	RM	LLL
R3	LUL	–	–	–
Matrix				
R1	RLL	RH	–	–
R2	RUL	M	–	–
R3	RUL	–	–	–

[a] See Table 2 for abbreviations.
[b] RN = *N*th retrieved image.

left lower lobe (bottom left), mediastinum (centre), as opposed to locations that fall between these locations, e.g. right side of the mediastinum. Again this is because matrices only store discrete descriptors of relative location and as such work best when the tumour is in an 'extreme' position. The MAP for the right middle lobe is low ($< 20\%$) for all methods because there were very few of these images within the dataset (5 of 50) meaning that other images could have been deemed to be more similar due to other features.

To achieve retrieval based on anatomical localisation of tumours, the CAPP graph similarity measurement algorithm needed to correctly match anatomical ROIs (vertices) across images (graphs). An incorrect anatomy matching would potentially lead to an incorrect tumour substitution, e.g. an incorrect match between the mediastinum and a lung would cause lung tumours to be considered as mediastinal tumours. Normally, anatomical matching could be done by matching labels assigned to the segmented anatomical ROIs. However, to not restrict our method to either general segmentation algorithms or structure-specific algorithms, we assumed in Section 2.2 that the segmented structures were not labelled. Therefore, in our case the anatomy assignment was entirely dependent upon the graph similarity calculation. We reduced the likelihood of an incorrect anatomical matching by representing all spatial relationships (even minor ones) between anatomical vertices, by creating edges between all pairs of anatomical vertices in our CAPP graphs, thus forming a fully connected subgraph (a complete anatomical subgraph).

The ability to retrieve images based on tumour localisation was also helped by our similarity measurement algorithm, which emphasised inter-class distances (Neuhaus et al., 2006). Calculating the similarity of multi-modality images with complete graphs favours substitution operations, particularly when the images have the same number of tumours. When CAPP graphs are used, all operations are used. As defined, insertion and deletion operations have a higher cost (see Eq. 8). This results in CAPP graphs having greater discriminatory power in cases with more than one tumour. As such, the CAPP graph had the highest precision when the graph dataset contained a wide variety of different structures, representing images with many tumours spread across the body.

In our graph representation each vertex or edge was treated as a feature vector. The graph similarity algorithm used these features (Section 4.3) and and the distance metric (Eq. 8) to measure the

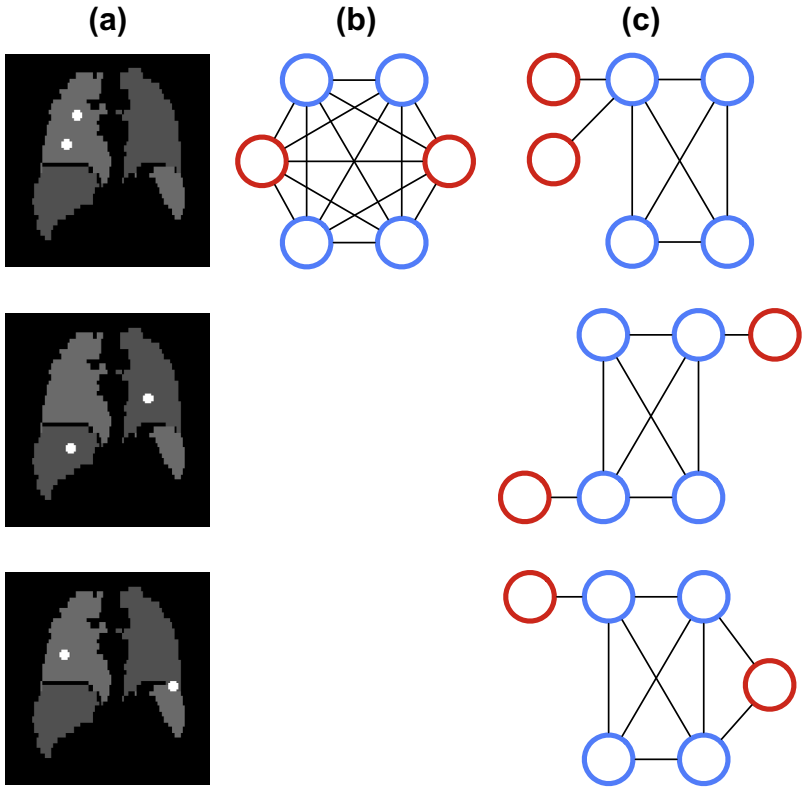

Fig. 8. Comparison of graphs representing images with the same number of tumours. Complete graphs always have the same structure, while CAPP graphs have more varied structures. Column (a): The image depicting four anatomical regions (grey ROIs) and two tumours (white ROIs). Column (b): The complete graph structure that represents all the images. Column (c): Distinct CAPP graph representations. Images in column (a) were derived from the OncoPET dataset (Tomei et al., 2010) used in our prior work (Kumar et al., 2012a).

level of similarity of different graph structures. For example, a small difference between two feature vectors indicated a low cost graph edit operation (implying a minor graph transformation). Table 5 shows that retrieval based purely on the CAPP graph structure (a graph with no image or relational features) had a lower MAP than both the CAPP and complete graphs with all features. We thus concluded that the higher retrieval precision was not purely the effect of the graph structure.

Table 5 also reveals that the precision of the CAPP graph was not greatly affected by a variation in the image feature sets. The normalised feature sets ensured that no individual feature would bias the metric because it had a higher range of values. The largest drop in retrieval precision occurred when the *ro* edge feature was removed from *F* because it was responsible for distinguishing between tumours in two different lobes of the same lung. For example, consider two single tumour images: one with a tumour in the left upper lobe and the other in the left lower lobe. Each of these images would have the same graph structure (vertex of the left lung connected to the tumour vertex). The difference in tumour location would thus be specified by the *ro* feature.

The reliance on segmentation is a limitation of our method, one not shared with other techniques like SIFT. Accurate segmentation is important for the extraction of ROI features (e.g., volume) and relationships (e.g., distance between ROIs). As such incorrect segmentation negatively impacts the similarity measurement. To this end in this work, we used semi-automatic segmentation techniques that could be manually adjusted (as described in Section 3.1). The refinement of the segmentation was mainly limited to the entry of 2 numerical parameters by typing or dragging sliders and an additional 2 mouse clicks for deleting extraneous (non-anatomy and non-tumour) regions. In the majority of the cases the segmentation was fine-tuned within 2 to 3 min.

When ROIs were detected by automatic segmentation the precision of our method was lower than when we used semi-automatic segmentation (Fig. 6, Tables 3 and 4). The precision of the methods that did not rely upon segmentation (SIFT and gray-level histograms) were unaffected. The results indicated that the CAPP graph method was more sensitive to accurate segmentation than the other segmentation reliant methods (complete graph and the relational matrix). This is because over-segmentation (the most common error in our experiments) distorts the proximity of the ROIs, thereby negatively impacting the edge pruning function (Eq. 2) used for CAPP graph construction. Despite the decrease in precision, the CAPP graph method still achieves an overall MAP that is comparable with the other retrieval methods. This indicated that existing automatic segmentation algorithms are already sufficiently reliable for use with the CAPP graph method.

The reliance on semi-automatic segmentation does not necessarily decrease the clinical usefulness of our method because the CAPP graph method had a higher overall MAP before the segmentation adjustments were applied (see Table 3). That is, the segmentation adjustments served to improve the overall retrieval precision. Integrating newer techniques that improve the reliability of automatic segmentation, such as the technique proposed by Kronman et al. (2013), will improve the retrieval precision when automatic segmentation techniques are used for ROI detection, thereby also improving the clinical usability of our retrieval method.

9. Conclusion

The CAPP graph representation enabled the retrieval of 3D multi-modality medical volumes based on the localisation of tumours in relation to anatomy. Our results show that the varied nature of

CAPP graph structures enabled higher levels of precision compared to several other retrieval methods. The CAPP graph is able to model tumour localisation within multi-modality images and it, together with our proposed similarity measurement algorithm, allowed modality-specific features to capture the complementary information inherent in multi-modal images.

Future work will include optimisation of the computational performance and accuracy of the similarity measurement of the CAPP graph. In particular, we will investigate: (i) the impact of adjusting the cost functions to emphasise edges between vertices in V_A and V_P; (ii) graph-based feature selection algorithms to enable the use of an optimal set of structural and image features; (iii) weighting graph substructures to emphasise tumours in specific locations of interest; and (iv) the optimal way of including state-of-the-art features (such as SIFT) as part of an extended graph feature set.

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.

Akakin, H., Gurcan, M., 2012. Content-based microscopic image retrieval system for multi-image queries. IEEE Trans. Inform. Technol. Biomed. 16, 758–769.

Akgül, C., Rubin, D., Napel, S., Beaulieu, C., Greenspan, H., Acar, B., 2011. Content-based image retrieval in radiology: current status and future directions. J. Digit. Imag. 24, 208–222.

Aksoy, S., Haralick, R.M., 2001. Feature normalization and likelihood-based similarity measures for image retrieval. Pattern Recogn. Lett. 22, 563–582.

Alajlan, N., Kamel, M.S., Freeman, G.H., 2008. Geometry-based image retrieval in binary image databases. IEEE Trans. Pattern Anal. Machine Intell. 30, 1003–1013.

Belhassen, S., Zaidi, H., 2010. A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET. Med. Phys. 37, 1309–1324.

Blodgett, T.M., Meltzer, C.C., Townsend, D.W., 2007. PET/CT: form and function. Radiology 242, 360–385.

Bradley, J., Thorstad, W.L., Mutic, S., Miller, T.R., Dehdashti, F., Siegel, B.A., Bosch, W., Bertrand, R.J., 2004. Impact of FDG-PET on radiation therapy volume delineation in non-small-cell lung cancer. Int. J. Radiat. Oncol. Biol. Phys. 59, 78–86.

Bugatti, P., Ponciano-Silva, M., Traina, A., Traina, A., Marques, P., 2009. Content-based retrieval of medical images: From context to perception. In: 22nd IEEE International Symposium on Computer-Based Medical Systems, pp. 1–8.

Bunke, H., 1999. Error correcting graph matching: on the influence of the underlying cost function. IEEE Trans. Pattern Anal. Machine Intell. 21, 917–922.

Bunke, H., Riesen, K., 2011. Recent advances in graph-based pattern recognition with applications in document analysis. Pattern Recog. 44, 1057–1067.

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 Res. 31, 1860–1861.

Conte, D., Foggia, P., Sansone, C., Vento, M., 2004. Thirty years of graph matching in pattern recognition. Int. J. Pattern Recogn. Artif. Intell. 18, 265–298.

Detterbeck, F.C., 2009. The new cancer staging system. Chest 136, 260–271.

S.B., Edge, Byrd, D.R., Compton, C., Frtiz, A.G., Greene, F.L., Trotti, A., 2010. AJCC Cancer Staging Manual, seventh ed. Springer, New York, NY.

Edge, S.B., Compton, C.C., 2010. The American joint committee on cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann. Surg. Oncol. 17, 1471–1474.

Fischer, B., Sauren, M., Guld, M.O., Deserno, T.M., 2008. Scene analysis with structural prototypes for content-based image retrieval in medicine. In: Medical Imaging 2008: Image Processing, SPIE, p. 69141X-9.

Foncubierta-Rodríguez, A., Poletti, P.A., Platon, A., Vargas, A., Müller, H., Depeursinge, A., 2013. Texture quantification in 4D dual energy CT for pulmonary embolism diagnosis. In: MICCAI Workshop on Medical Content-Based Retrieval for Clinical Decision Support 2012, pp. 45–56.

Ghosh, P., Antani, S., Long, L., Thoma, G., 2011. Review of medical image retrieval systems and future directions. In: 24th International Symposium on Computer-Based Medical Systems, pp. 1–6.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybernet. 3, 610–621.

Hart, P.E., Nilsson, N.J., Raphael, B., 1968. A formal basis for the heuristic determination of minimum cost paths. IEEE Trans. Syst. Sci. Cybernet. 4, 100–107.

Hoàng, N., Gouet-Brunet, V., Rukoz, M., Manouvrier, M., 2010. Embedding spatial information into image content description for scene retrieval. Pattern Recogn. 43, 3013–3024.

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.

Huang, M., Yang, W., Yu, M., Lu, Z., Feng, Q., Chen, W., 2012. Retrieval of brain tumors with region-specific bag-of-visual-words representations in contrast-enhanced MRI images. Comput. Math. Methods Med. 2012, 17.

Iakovidis, D.K., Pelekis, N., Kotsifakos, E.E., Kopanakis, I., Karanikas, H., Theodoridis, Y., 2009. A pattern similarity scheme for medical image retrieval. IEEE Trans. Inform. Technol. Biomed. 13, 442–450.

Jaworska, T., Kacprzyk, J., Marín, N., Zadrożny, S., 2010. On dealing with imprecise information in a content based image retrieval system. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (Eds.), Computational Intelligence for Knowledge-Based Systems Design, Lecture Notes in Computer Science, vol. 6178. Springer, Berlin Heidelberg, pp. 149–158.

Jouili, S., Tabbone, S., 2012. Hypergraph-based image retrieval for graph-based representation. Pattern Recogn. 45, 4054–4068.

Kim, J., Cai, W., Feng, D., 2009. Bridging the feature gaps for retrieval of multi-dimensional images. Int. J. Healthcare Inform. Syst. Inform. 4, 34–46.

Kronman, A., Joskowicz, L., Sosna, J., 2013. Gradient-based correction of segmentation boundary inaccuracies in medical CT images. Int. J. Comput. Assist. Radiol. Surg. 8 (Suppl. 1), S49–51.

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. In: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 390–393.

Kumar, A., Kim, J., Feng, D., Fulham, M., 2012a. Graph-based retrieval of multi-modality medical images: A comparison of representations using simulated images. In: 25th International Symposium on Computer-Based Medical Systems, pp. 1–6.

Kumar, A., Kim, J., Wen, L., Feng, D., 2012b. A graph-based approach to the retrieval of volumetric PET-CT images. In: 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5408–5411.

Lee, D.J., Antani, S., Chang, Y., Gledhill, K., Rodney Long, L., Christensen, P., 2009. CBIR of spine X-ray images on inter-vertebral disc space and shape profiles using feature ranking and voting consensus. Data Knowl. Eng. 68, 1359–1369.

Lister, T.A., Crowther, D., Sutcliffe, S.B., Glatstein, E., Canellos, G.P., Young, R.C., Rosenberg, S.A., Coltman, C.A., Tubiana, M., 1989. Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds meeting. J. Clin. Oncol. 7, 1630–1636.

Llados, J., Marti, E., Villanueva, J., 2001. Symbol recognition by error-tolerant subgraph matching between region adjacency graphs. IEEE Trans. Pattern Anal. Machine Intell. 23, 1137–1143.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60, 91–110.

Müller, H., Deserno, T., 2011. Content-based medical image retrieval. In: Deserno, T.M. (Ed.), Biomedical Image Processing, Biological and Medical Physics, Biomedical Engineering. Springer, Berlin Heidelberg, pp. 471–494.

Müller, H., Kalpathy-Cramer, J., Caputo, B., Syeda-Mahmood, T., Wang, F., 2009. Overview of the First Workshop on Medical Content-Based Retrieval for Clinical Decision Support at MICCAI 2009. Lecture Notes in Computer Science, vol. 5853. Springer, Berlin/Heidelberg.

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

Müller, H., Rosset, A., Garcia, A., Vallee, J.P., Geissbuhler, A., 2005. Informatics in radiology (infoRAD): benefits of content-based visual data access in radiology. Radiographics 25, 849–858.

Müller, H., Xin, Z., Depeursinge, A., Pitkanen, M., Iavindrasana, J., Geissbuhler, A., 2007. Medical Visual Information Retrieval: state of the art and challenges ahead. In: IEEE International Conference on Multimedia and Expo, pp. 683–686.

Napel, S.A., Beaulieu, C.F., Rodriguez, C., Cui, J., Xu, J., Gupta, A., Korenblum, D., Greenspan, H., Ma, Y., Rubin, D.L., 2010. Automated retrieval of CT images of liver lesions on the basis of image similarity: method and preliminary results. Radiology 256, 243–252.

Neuhaus, M., Riesen, K., Bunke, H., 2006. Fast suboptimal algorithms for the computation of graph edit distance. Struct. Syntact. Stat. Pattern Recogn. 4109, 163–172.

Petrakis, E.G.M., 2002. Design and evaluation of spatial similarity approaches for image retrieval. Image Vis. Comput. 20, 59–76.

Petrakis, E.G.M., Faloutsos, C., Lin, K.I., 2002. ImageMap: an image indexing method based on spatial similarity. IEEE Trans. Knowl. Data Eng. 14, 979–987.

Quellec, G., Lamard, M., Bekri, L., Cazuguel, G., Roux, C., Cochener, B., 2010a. Medical case retrieval from a committee of decision trees. IEEE Trans. Inform. Technol. Biomed. 14, 1227–1235.

Quellec, G., Lamard, M., Cazuguel, G., Cochener, B., Roux, C., 2010b. Wavelet optimization for content-based image retrieval in medical databases. Med. Image Anal. 14, 227–241.

Quellec, G., Lamard, M., Cazuguel, G., Roux, C., Cochener, B., 2011. Case retrieval in medical databases by fusing heterogeneous information. IEEE Trans. Med. Imag. 30, 108–118.

Shyu, C.R., Brodley, C.E., Kak, A.C., Kosaka, A., Aisen, A.M., Broderick, L.S., 1999. ASSERT: a physician-in-the-loop content-based retrieval system for HRCT image databases. Comput. Vis. Image Understand. 75, 111–132.

Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R., 2000. Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Machine Intell. 22, 1349–1380.

Song, Y., Cai, W., Eberl, S., Fulham, M.J., Feng, D., 2010. A content-based image retrieval framework for multi-modality lung images. In: IEEE 23rd International Symposium on Computer-Based Medical Systems, pp. 285–290.

Song, Y., Cai, W., Eberl, S., Fulham, M.J., Feng, D., 2011. Thoracic image case retrieval with spatial and contextual information. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1885–1888.

Tomei, S., Reilhac, A., Visvikis, D., Boussion, N., Odet, C., Giammarile, F., Lartizien, C., 2010. OncoPET_DB: a freely distributed database of realistic simulated whole body 18F-FDG PET images for oncology. IEEE Trans. Nucl. Sci. 57 (Part 1), 246–255.

Townsend, D.W., Beyer, T., Blodgett, T.M., 2003. PET/CT scanners: a hardware approach to image fusion. Semin. Nucl. Med. 33, 193–204.

Wadell, H., 1935. Volume, shape, and roundness of quartz particles. J. Geol. 43, 250–280.

Wang, S., Liu, D., Gu, Fangming, Feng Yang, H.L., 2012. Similar matching for images with complex spatial relations. J. Comput. Inform. Syst. 8, 8727–8734.

Xu, X., Lee, D.J., Antani, S., Long, L.R., 2008. A spine X-ray image retrieval system using partial shape matching. IEEE Trans. Inform. Technol. Biomed. 12, 100–108.

Zhou, X., Stern, R., Müller, H., 2012. Case-based fracture image retrieval. Int. J. Comput. Assist. Radiol. Surg. 7, 401–411.