

COPYRIGHT AND USE OF THIS THESIS

This thesis must be used in accordance with the provisions of the Copyright Act 1968.

Reproduction of material protected by copyright may be an infringement of copyright and copyright owners may be entitled to take legal action against persons who infringe their copyright.

Section 51 (2) of the Copyright Act permits an authorized officer of a university library or archives to provide a copy (by communication or otherwise) of an unpublished thesis kept in the library or archives, to a person who satisfies the authorized officer that he or she requires the reproduction for the purposes of research or study.

The Copyright Act grants the creator of a work a number of moral rights, specifically the right of attribution, the right against false attribution and the right of integrity.

You may infringe the author's moral rights if you:

- fail to acknowledge the author of this thesis if you quote sections from the work
- attribute this thesis to another author
- subject this thesis to derogatory treatment which may prejudice the author's reputation

For further information contact the University's Director of Copyright Services

sydney.edu.au/copyright

A GRAPH-BASED APPROACH FOR THE RETRIEVAL OF MULTI-MODALITY MEDICAL IMAGES



A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy.

Faculty of Engineering & Information Technologies The University of Sydney

> Ashnil Kumar March 2013

© Copyright by Ashnil Kumar 2013 All Rights Reserved

Abstract

Advances in sensors and imaging technologies are contributing to rapidly expanding data repositories that contain interrelated information from different modalities. The extraction and visualisation of knowledge from these repositories is a major challenge in the modern, digital world. In the medical domain, images are routinely acquired for a variety of tasks, including diagnosis and patient monitoring. Advances in imaging technologies have resulted in devices capable of acquiring images in multiple dimensions (volumetric and dynamic) as well as from multiple modalities. One example of a widely used volumetric and multimodality image is combined positron emission tomography and computed tomography (PET-CT), which presents physicians with complementary functional and anatomical features and spatial relationships. In clinical practice, PET-CT imaging has already proven its ability to improve cancer diagnosis, localisation, and staging compared to its single-modality counterparts.

The clinical benefits provided by medical imaging have spurred increases in the data volume acquired in clinical environments. As such, massive medical imaging collections offer the opportunity for search-based applications in evidence-based diagnosis, physician training, and biomedical research. However, conventional search techniques that operate upon manually assigned textual annotations are not feasible for the volume of data acquired in modern hospitals. Qualitative text descriptions are also limited in their capacity to quantitatively describe the rich information inherent in medical images.

Content-based image retrieval (CBIR) is an image search technique that utilises visual features as search criteria. CBIR has already demonstrated benefits for evidence-based diagnosis, physician training, and biomedical research by allowing clinical staff to consider relevant knowledge from retrieved cases. The majority of medical CBIR research has focused on single modality medical images leaving a clear deficiency in the retrieval of multi-modality images. In particular, images like PET-CT offer the ability for retrieval based upon the relationships between regions in different modalities, such as the location of tumour features (from PET) in relation to organ features (from CT). The challenge of multi-modality image retrieval for cancer patients lies in representing these complementary geometric and topologic attributes between tumours and organs. A secondary challenge lies in the human aspect of retrieval – effectively communicating the retrieved results to users and facilitating a better understanding of the similarity between the query and retrieved multi-modality images.

As such, in this thesis we propose a new graph representation for multimodality images. Our representation preserves the spatial relationships between modalities by emphasising the inherent characteristics of these images that are used for disease staging and classification. This is done by structurally constraining the graph based on image features, e.g., spatial proximity of tumours and organs. We also present a similarity matching algorithm that accounts for different feature sets for graph elements from different imaging modalities. Our approach prioritises 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 localisation.

We also propose a method for defining user interfaces (UIs) that enable effective human interpretation of retrieved multi-modality images. A set of visualisation and interaction requirements based on the characteristics of PET-CT images were used to implement a CBIR UI. The UI visualised multiple views of a single image, displayed abstractions of image data, and provided access to supplementary non-image data. We also defined interactions for visually indicating the similarities between 3D regions, e.g., similar tumours.

We evaluated our retrieval methodology on three data sets: simulated 2D liver shape images, simulated 3D lymphoma images, and clinical PET-CT volumes. Our results demonstrated that our method achieved a high retrieval precision, especially in three common scenarios: (1) retrieving images with multiple tumours spread across multiple organs, (2) retrieving images with multiple shape distortions, and (3) retrieving images from a data set with large anatomical and tumour variations. In particular, our algorithm retrieved images on the basis of tumour location within organs. The evaluation of our proposed UI design by user surveys revealed that it improved the ability of users to interpret and understand the similarity between retrieved PET-CT images. The work in this thesis advances the state-of-the-art by enabling a novel approach for the retrieval of multi-modality medical images.

Acknowledgements

"It's bigger on the inside!"

Those words, exclaimed by many of The Doctor's various companions¹, echo in my mind as I sit here reflecting upon the entirety of my PhD. My experience during this degree program was indeed bigger than any of my expectations. The challenges were more difficult than I had thought, the community larger than I had imagined, the highs (and the lows) more frightening than I would care to admit.

Just as each companion assists The Doctor and shapes his character in their journeys through time and space, so have a number of people stood by me during my quest. Without such stalwart friends and colleagues, I would have been left beleaguered by the roadside and it is thanks to them that I stand near the end of my journey. Words cannot begin to express the depth of my gratitude but I will do my best anyway.

My supervisors, David Feng, Tom Cai, and Jinman Kim, welcomed a very naïve person into their research group. He was convinced he had a brilliant idea, one that would change the world. They gave me the freedom to pursue my own research while protecting me from the cliffs I ventured recklessly towards. Their patience and guidance was irreplaceable. Their advice: invaluable. I may have abused Jinman's open door policy, accosting him many times about minor details surrounding my work. He never complained and always stood ready to give me the resources and support I needed. They also encouraged me to tackle a significant research problem. Their encouragement and confidence in my talents helped me to secure various awards during my candidature, including the Microsoft Research Asia Fellowship, the Japan-East Asia Network of Exchange for Students and Youths Scholarship, and the CARS 2012 - EuroPACS Poster Award.

Lingfeng Wen, Stefan Eberl, and Michael Fulham at the Royal Prince Alfred

 $^{^1\}mathrm{Characters}$ on the excellent BBC TV show *Doctor Who*.

Hospital in Sydney were always available when I needed input about the clinical relevance of my work. I always skirted close to publication deadlines and they never complained about the rushed schedule. In particular, I am extremely grateful to Michael's "harsh justice", which was instrumental in enhancing my work.

I have had the honour and the opportunity of supervising and collaborating with several Masters students during my PhD: Jerry Ling, Daiki Haraguchi, and Lei Bi. They each contributed their expertise to my work and also helped me to write code. Without their efforts I would have less to show. I am also grateful for the opportunity to work alongside Ben Niehaus and Falk Nette, two visiting students from Germany, as well as Youn Hyun Jung.

I should not forget my fellow compatriots in this journey towards the PhD: Jan Chan, Peter Budd, and Liviu Constantinescu. Jan, Pete, and Liviu set a high standard and I have done my best to match their enthusiasm and diligent work ethic. I am eternally grateful to Liviu for always being there to listen to my concerns and for sharing his experiences with me. Our collaboration on the INVOLVE2 project and our various trips abroad were some of the most enjoyable parts of my PhD. I hope I will be able to be as generous with my time and expertise towards future generations of students.

During my PhD I had the good fortune of being able to travel to other institutions. These visits enriched my own knowledge, enhanced my world views, and gave me the opportunity to sample some exotic cuisine. I would like to thank Takao Nishizeki, Takehiro Ito, and Kei Uchizawa at Tohoku University, Zheru Chi and Guangjiang Tian at Hong Kong Polytechnic University, and Wenwu Zhu and Wen Sun from Microsoft Research Asia. I would also like to thank John Warren from Microsoft for his efforts in securing my trip to MSRA.

I must also acknowledge the contributions of my lunch group, especially Liviu, Pete, Emma, Andrew, Kaz, and Miranda. My somewhat disturbing love of fast food and hot sauce never once frightened them and our lunch breaks were welcome distractions from my ongoing work.

The support of my family was integral to my mental well-being. I would have not survived this journey without the love, affection, and food they provided. Thank you very much to my father, Anil, my mother, Renuka, and my little brother, Vinil. Thank you also to my aunt and uncle, Sadhana and André, and my cousin, Alex, for their constant encouragement. I am incredibly grateful to my martial arts teachers, Gillian Booth and Frances Haynes, for helping me become a more physically active and healthier person. Their teachings helped me to discover the inner strength that always eluded me. I am also thankful to my wonderful dancing teachers, especially Katy and Rosie. Their encouragement has done wonders for my self-confidence and sociability. I have applied their lessons to every part of my daily life and I am happier for it.

I have a lot of friends. They all know that I can become very obsessed with my research work at the expense of other aspects of my life. I am incredibly grateful for their concerted efforts to ensure that I did not work myself to an early grave. They provided me with outlets for my stress and helped me to become a less worried and more well-adjusted person. Thank you to Liviu (again), Pete (again), Wei, Ryan, Jason, Zach, Bogdan, Dane, Pete (the other one), Andrew (again), Sam, Mark, Sean, Annabel, Adam, and Pete (the other, other one).

I am sure I have missed naming someone who has been critical to my success. They are omitted not out of malice but because of my poor, failing memory. To them, and all others mentioned here, I say: thank you.

Contents

Abstract				
A	cknov	wledgements	vii	
\mathbf{Li}	st of	Tables	xv	
\mathbf{Li}	st of	Figures	xvii	
\mathbf{Li}	st of	Algorithms	xviii	
\mathbf{Li}	st of	Publications	xxi	
1	Intr	roduction	1	
	1.1	Motivation	1	
	1.2	Aims and Objectives	6	
	1.3	Contributions of this Thesis	6	
	1.4	Thesis Structure	8	
	1.5	Key Terms	9	
2	Mee	dical Imaging and Image Processing	12	
	2.1	Definitions	12	
	2.2	Single Modality Medical Imaging	14	
		2.2.1 X-ray Imaging	14	
		2.2.2 Computed Tomography	15	
		2.2.3 Positron Emission Tomography	16	
		2.2.4 Magnetic Resonance Imaging	17	
	2.3	Multi-Modality Medical Imaging	18	
		2.3.1 PET-CT	19	
		2.3.2 PET-MR	20	

	2.4	Image	Processing	20
		2.4.1	Segmentation	20
		2.4.2	Registration	21
		2.4.3	Visualisation	23
3	Cor	ntent-E	Based Image Retrieval	24
	3.1	Measu	ring Image Similarity	24
	3.2	Conte	nt-Based Image Retrieval	25
		3.2.1	Features and Representations	29
		3.2.2	Similarity Measurement	31
		3.2.3	Display and Feedback	32
	3.3	Medic	al Content-Based Image Retrieval	33
		3.3.1	2D Medical Image Retrieval	35
		3.3.2	3D+ Medical Image Retrieval	38
		3.3.3	Retrieval Enhancement Using Non-Image Data	40
		3.3.4	Retrieval from Diverse Data Sets	42
		3.3.5	Multiple Images and Modalities	44
	3.4	Summ	ary of Gaps	45
		3.4.1	Multi-dimensional Image Processing and Feature Extraction	47
		3.4.2	Interpretation and Visualisation	49
		3.4.3	Inclusion of Clinical Context	50
		3.4.4	Data Sets for Evaluation	51
4	AC	Graph	Primer	52
	4.1	Defini	tions	52
	4.2	Repre	senting Structures with Graphs	54
	4.3	Measu	ring Graph Similarity	57
		4.3.1	Graph Edit Distance	57
		4.3.2	Graph Kernels	61
5	CB	IR Des	sign	64
	5.1	Overv	iew	64
		5.1.1	Assumptions	66
	5.2	Frame	ework	67
		5.2.1	Data Set Indexing	67
		5.2.2	Querying	69

 6 Multi-Modality Image Retrieval 6.1 Overview. 6.2 Multi-modality Graph Definitions 6.2.1 Complete Graph 6.2.2 Complete-Anatomy Proximal-Pathology Graph 6.3 Graph Attributes and Image Features 6.3.1 Types of Features 6.3.2 Image Features 6.3.3 Feature Set Justification 6.4 Graph Feature Normalisation 6.4.1 Measurements 6.4.2 Angular Values 6.4.3 Point Sets 6.5 Retrieval Procedure 6.6 Graph Comparisons 6.6.1 Determining Operation Costs 6.6.2 Beam Search Adaptation 6.6.3 Aside: An Inferior Approach 7 Evaluation of Retrieval Method 7.1 Data Sets 7.1.2 OncoPET: Simulated 3D Lymphoma Images 7.1.3 Clinical PET-CT Images 7.2 Evaluating the Relevance of Retrieved Images 7.3.3 Clinical PET-CT 7.3.3 Clinical PET-CT 7.3.4 Computational Performance 	5.3	Design	Justification	69
6.1 Overview 6.2 Multi-modality Graph Definitions 6.2.1 Complete Graph 6.2.2 Complete Graph 6.3.3 Graph Attributes and Image Features 6.3.4 Graph Attributes and Image Features 6.3.5 Image Features 6.3.6 Graph Attributes and Image Features 6.3.1 Types of Features 6.3.2 Image Features 6.3.3 Feature Set Justification 6.4.3 Feature Normalisation 6.4.4 Angular Values 6.4.5 Retrieval Procedure 6.4.6 Graph Comparisons 6.5 Retrieval Procedure 6.6.1 Determining Operation Costs 6.6.2 Beam Search Adaptation 6.6.3 Aside: An Inferior Approach 7.1 Data Sets 7.1.2 OncoPET: Simulated 3D Lymphoma Images 7.1.3 Clinical PET-CT Images 7.2.1 Determining the Relevance of Retrieved Images 7.2.2 Evaluating the Computational Performance 7.3.3 Clinical PET-CT 7.3.4 Computational Performance	Mul	ti-Moo	dality Image Retrieval	72
 6.2 Multi-modality Graph Definitions	6.1	Overvi	ew	72
6.2.1 Complete Graph 6.2.2 Complete-Anatomy Proximal-Pathology Graph 6.3 Graph Attributes and Image Features 6.3.1 Types of Features 6.3.2 Image Features 6.3.3 Feature Set Justification 6.4 Graph Feature Normalisation 6.4.1 Measurements 6.4.2 Angular Values 6.4.3 Point Sets 6.5 Retrieval Procedure 6.6 Graph Comparisons 6.6.1 Determining Operation Costs 6.6.2 Beam Search Adaptation 6.6.3 Aside: An Inferior Approach 7.1 Dimulated 2D Liver Shapes 7.1.1 Simulated 2D Liver Shapes 7.1.2 OncoPET: Simulated 3D Lymphoma Images 7.2 Experimental Procedure 7.2.1 Determining the Relevance of Retrieved Images 7.3.1 Liver Shapes 7.3.2 OncoPET 7.3.3 Clinical PET-CT 7.3.4 Computational Performance	6.2	Multi-	modality Graph Definitions	74
 6.2.2 Complete-Anatomy Proximal-Pathology Graph 6.3 Graph Attributes and Image Features 6.3.1 Types of Features 6.3.2 Image Features 6.3.3 Feature Set Justification 6.4 Graph Feature Normalisation 6.4 Graph Feature Normalisation 6.4.1 Measurements 6.4.2 Angular Values 6.4.3 Point Sets 6.5 Retrieval Procedure 6.6 Graph Comparisons 6.6.1 Determining Operation Costs 6.6.2 Beam Search Adaptation 6.6.3 Aside: An Inferior Approach 6.6.3 Aside: An Inferior Approach 7 Evaluation of Retrieval Method 7.1 Data Sets 7.1.1 Simulated 2D Liver Shapes 7.1.2 OncoPET: Simulated 3D Lymphoma Images 7.2 Experimental Procedure 7.3 Clinical PET-CT Images 7.3.1 Liver Shapes 7.3.2 OncoPET 7.3.3 Clinical PET-CT 7.3.4 Computational Performance 		6.2.1	Complete Graph	74
 6.3 Graph Attributes and Image Features		6.2.2	Complete-Anatomy Proximal-Pathology Graph	75
6.3.1 Types of Features 6.3.2 Image Features 6.3.3 Feature Set Justification 6.4 Graph Feature Normalisation 6.4.1 Measurements 6.4.2 Angular Values 6.4.3 Point Sets 6.4.3 Point Sets 6.4.3 Point Sets 6.4.4 Procedure 6.5 Retrieval Procedure 6.6 Graph Comparisons 6.6.1 Determining Operation Costs 6.6.2 Beam Search Adaptation 6.6.3 Aside: An Inferior Approach 7.1 Data Sets 7.1.1 Simulated 2D Liver Shapes 7.1.2 OncoPET: Simulated 3D Lymphoma Images 7.1.3 Clinical PET-CT Images 7.2.1 Determining the Relevance of Retrieved Images 7.2.2 Evaluating the Computational Performance 7.3.1 Liver Shapes 7.3.2 OncoPET 7.3.3 Clinical PET-CT 7.3.4 Computational Performance	6.3	Graph	Attributes and Image Features	77
6.3.2 Image Features 6.3.3 Feature Set Justification 6.4 Graph Feature Normalisation 6.4.1 Measurements 6.4.2 Angular Values 6.4.3 Point Sets 6.4.3 Point Sets 6.4.3 Point Sets 6.4.3 Point Sets 6.4.4 Procedure 6.4.5 Retrieval Procedure 6.6 Graph Comparisons 6.6.6 Graph Comparisons 6.6.7 Beam Search Adaptation 6.6.8 Aside: An Inferior Approach 6.6.9 Beam Search Adaptation 6.6.1 Determining Operation Costs 6.6.2 Beam Search Adaptation 6.6.3 Aside: An Inferior Approach 7.1 Data Sets 7.1.1 Simulated 2D Liver Shapes 7.1.2 OncoPET: Simulated 3D Lymphoma Images 7.1.3 Clinical PET-CT Images 7.2.1 Determining the Relevance of Retrieved Images 7.2.2 Evaluating the Computational Performance 7.3.4 Computational Performance 7.3.4		6.3.1	Types of Features	77
 6.3.3 Feature Set Justification 6.4 Graph Feature Normalisation 6.4.1 Measurements 6.4.2 Angular Values 6.4.3 Point Sets 6.4.3 Point Sets 6.5 Retrieval Procedure 6.6 Graph Comparisons 6.6.1 Determining Operation Costs 6.6.2 Beam Search Adaptation 6.6.3 Aside: An Inferior Approach 7 Evaluation of Retrieval Method 7.1 Data Sets 7.1.1 Simulated 2D Liver Shapes 7.1.2 OncoPET: Simulated 3D Lymphoma Images 7.2 Experimental Procedure 7.2.1 Determining the Relevance of Retrieved Images 7.2.2 Evaluating the Computational Performance 7.3 Results 7.3.4 Computational Performance 		6.3.2	Image Features	78
 6.4 Graph Feature Normalisation		6.3.3	Feature Set Justification	81
6.4.1 Measurements 6.4.2 Angular Values 6.4.3 Point Sets 6.4.3 Point Sets 6.5 Retrieval Procedure 6.6 Graph Comparisons 6.6 Graph Comparisons 6.6.1 Determining Operation Costs 6.6.2 Beam Search Adaptation 6.6.3 Aside: An Inferior Approach 6.6.3 Aside: An Inferior Approach 7 Evaluation of Retrieval Method 7.1 Data Sets 7.1.1 Simulated 2D Liver Shapes 7.1.2 OncoPET: Simulated 3D Lymphoma Images 7.1.3 Clinical PET-CT Images 7.2 Experimental Procedure 7.2.1 Determining the Relevance of Retrieved Images 7.2.2 Evaluating the Computational Performance 7.3 Results 7.3.1 Liver Shapes 7.3.2 OncoPET 7.3.3 Clinical PET-CT 7.3.4 Computational Performance	6.4	Graph	Feature Normalisation	82
6.4.2 Angular Values 6.4.3 Point Sets 6.5 Retrieval Procedure 6.6 Graph Comparisons 6.6 Graph Comparisons 6.6.1 Determining Operation Costs 6.6.2 Beam Search Adaptation 6.6.3 Aside: An Inferior Approach 7 Evaluation of Retrieval Method 7.1 Data Sets 7.1.1 Simulated 2D Liver Shapes 7.1.2 OncoPET: Simulated 3D Lymphoma Images 7.1.3 Clinical PET-CT Images 7.2 Experimental Procedure 7.2.1 Determining the Relevance of Retrieved Images 7.2.2 Evaluating the Computational Performance 7.3.1 Liver Shapes 7.3.2 OncoPET 7.3.3 Clinical PET-CT 7.3.4 Computational Performance		6.4.1	Measurements	83
6.4.3 Point Sets 6.5 Retrieval Procedure 6.6 Graph Comparisons 6.6.1 Determining Operation Costs 6.6.1 Determining Operation Costs 6.6.2 Beam Search Adaptation 6.6.3 Aside: An Inferior Approach 7 Evaluation of Retrieval Method 7.1 Data Sets 7.1.1 Simulated 2D Liver Shapes 7.1.2 OncoPET: Simulated 3D Lymphoma Images 7.1.3 Clinical PET-CT Images 7.2 Experimental Procedure 7.2.1 Determining the Relevance of Retrieved Images 7.2.2 Evaluating the Computational Performance 7.3.1 Liver Shapes 7.3.2 OncoPET 7.3.3 Clinical PET-CT 7.3.4 Computational Performance		6.4.2	Angular Values	83
 6.5 Retrieval Procedure		6.4.3	Point Sets	84
6.6 Graph Comparisons	6.5	Retriev	val Procedure	84
6.6.1 Determining Operation Costs . 6.6.2 Beam Search Adaptation . 6.6.3 Aside: An Inferior Approach . 7 Evaluation of Retrieval Method 7.1 Data Sets . 7.1.1 Simulated 2D Liver Shapes . 7.1.2 OncoPET: Simulated 3D Lymphoma Images . 7.1.3 Clinical PET-CT Images . 7.2 Experimental Procedure . 7.2.1 Determining the Relevance of Retrieved Images . 7.2.2 Evaluating the Computational Performance . 7.3.1 Liver Shapes . 7.3.2 OncoPET . 7.3.3 Clinical PET-CT . 7.3.4 Computational Performance .	6.6	Graph	Comparisons	85
6.6.2 Beam Search Adaptation . 6.6.3 Aside: An Inferior Approach . 7 Evaluation of Retrieval Method 7.1 Data Sets . 7.1.1 Simulated 2D Liver Shapes . 7.1.2 OncoPET: Simulated 3D Lymphoma Images . 7.1.3 Clinical PET-CT Images . 7.2 Experimental Procedure . 7.2.1 Determining the Relevance of Retrieved Images . 7.2.2 Evaluating the Computational Performance . 7.3 Results . 7.3.1 Liver Shapes . 7.3.2 OncoPET . 7.3.3 Clinical PET-CT . 7.3.4 Computational Performance . 7.4 Discursion .		6.6.1	Determining Operation Costs	88
6.6.3 Aside: An Inferior Approach 7 Evaluation of Retrieval Method 7.1 Data Sets 7.1.1 Simulated 2D Liver Shapes 7.1.2 OncoPET: Simulated 3D Lymphoma Images 7.1.3 Clinical PET-CT Images 7.1 Determining the Relevance of Retrieved Images 7.2 Experimental Procedure 7.2.1 Determining the Relevance of Retrieved Images 7.2.2 Evaluating the Computational Performance 7.3.1 Liver Shapes 7.3.2 OncoPET 7.3.3 Clinical PET-CT 7.3.4 Computational Performance		6.6.2	Beam Search Adaptation	90
 7 Evaluation of Retrieval Method 7.1 Data Sets		6.6.3	Aside: An Inferior Approach	93
 7.1 Data Sets	Eva	luation	of Retrieval Method	94
 7.1.1 Simulated 2D Liver Shapes	7.1	Data S	Sets	94
 7.1.2 OncoPET: Simulated 3D Lymphoma Images		7.1.1	Simulated 2D Liver Shapes	95
 7.1.3 Clinical PET-CT Images		7.1.2	OncoPET: Simulated 3D Lymphoma Images	97
 7.2 Experimental Procedure		7.1.3	Clinical PET-CT Images	98
 7.2.1 Determining the Relevance of Retrieved Images 7.2.2 Evaluating the Computational Performance 7.3 Results	7.2	Experi	mental Procedure	100
7.2.2 Evaluating the Computational Performance 7.3 Results 7.3.1 Liver Shapes 7.3.2 OncoPET 7.3.3 Clinical PET-CT 7.3.4 Computational Performance		7.2.1	Determining the Relevance of Retrieved Images	100
7.3 Results 7.3.1 Liver Shapes 7.3.1 Liver Shapes 7.3.2 OncoPET 7.3.2 OncoPET 7.3.3 Clinical PET-CT 7.3.3 Clinical PET-CT 7.3.4 Computational Performance 7.4 Discussion		7.2.2	Evaluating the Computational Performance	102
7.3.1 Liver Shapes 7.3.2 OncoPET 7.3.3 Clinical PET-CT 7.3.4 Computational Performance	7.3	Result	8	103
 7.3.2 OncoPET		7.3.1	Liver Shapes	103
 7.3.3 Clinical PET-CT		7.3.2	OncoPET	105
7.3.4 Computational Performance		7.3.3	Clinical PET-CT	106
7.4 Discussion		7.3.4	Computational Performance	107
$(.4 DISCUSSIOII \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots $	7.4	Discus	sion	112
7.5 Summary	7.5	Summ	ary	121

8	Gra	ph-Bas	sed Retrieval Interpretation	123
	8.1	Integra	ation with the CBIR Framework	. 123
	8.2	UI and	d Visualisation Requirements	. 124
	8.3	User In	nterface Design	. 125
		8.3.1	Visualising Multiple Views of Volumetric Images	. 127
		8.3.2	Abstraction Visualisation	. 129
	8.4	User In	nteractions	. 130
		8.4.1	Changing Data Presentation	. 131
		8.4.2	Query Refinement	. 132
		8.4.3	Visual Indication of Similarity	. 132
9	Eva	luation	n of Retrieval User Interface	138
	9.1	Evalua	ation Procedure	. 138
		9.1.1	Materials	. 138
		9.1.2	Survey Procedure	. 139
		9.1.3	Other Evaluation $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$. 142
	9.2	Result	s	. 142
		9.2.1	Survey Responses	. 142
		9.2.2	Participant Comments	. 142
		9.2.3	Performance Evaluation	. 143
	9.3	Discus	sion \ldots	. 144
10	Disc	cussion	1	152
	10.1	Capab	ilities	. 152
		10.1.1	Structural Extensibility	. 152
		10.1.2	Modality Extensibility	. 153
		10.1.3	Feature Extensibility	. 153
		10.1.4	Full Representation - Complete Graphs	. 154
		10.1.5	Structural Representation - CAPP Graphs $\ldots \ldots \ldots$. 154
	10.2	Future	e Research Directions	. 155
		10.2.1	Probabilistic Graph Construction	. 155
		10.2.2	Including Temporal Information	. 156
		10.2.3	Improved Segmentation and Registration	. 156
		10.2.4	Optimal Feature Selection	. 157
		10.2.5	Reduction in Computational Complexity	. 157
		10.2.6	Designing an A [*] Heuristic Function $\ldots \ldots \ldots \ldots$. 158

10.2.7 Vector Space Embedding $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots $	158
10.2.8 Hardware-Driven Graph Matching	159
10.2.9 Data Mining and Classification $\ldots \ldots \ldots \ldots \ldots \ldots \ldots $	159
$10.2.10$ Improved Visualisation Techniques $\ldots \ldots \ldots \ldots \ldots $	159
$10.2.11$ Integration of Visual Analytics $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots $	160
11 Conclusions1Appendices1	.61 .63
A Graph Feature Normalisation 1	.64
A.1 Index Normalisation $\ldots \ldots 1$	164
A.2 Calculating Feature Distributions	165
A.3 Graph Normalisation	167
References 1	.69

List of Tables

3.1	Medical CBIR Studies Divided by Data Types	35
7.1	Summary of Data Sets	95
7.2	OncoPET Retrieval Localisation	06
7.3	MAP By Tumour Location	08
7.4	Feature Contributions	08
7.5	Clinical Retrieval Localisation	09
7.6	Graph Comparison Time Summary (Images with 5 Tumours) 1	12
9.1	User Survey Statements	41
9.2	Survey Response Summary	44
9.3	Resource Usage	46
9.4	Task Timing	46

List of Figures

2.1	The pixel and spatial resolution of images	13
2.2	The region of interest in an image.	14
2.3	Three CT tomograms (slices) from the same volume. \ldots .	15
2.4	Slices from different planes of the same PET volume. \ldots .	17
2.5	T2 weighted brain MR image slices	18
2.6	A PET-CT image.	19
2.7	Segmenting a chest CT image	21
2.8	Registering two CT images.	22
3.1	Renaissance art	25
4.1	A graph with five vertices	53
4.2	Two subgraphs of Figure 4.1.	53
4.3	Isomorphism and subgraph isomorphism	54
4.4	Constructing a graph representation of a pattern. \ldots	56
4.5	Graph edit distance search tree	59
5.1	Proposed CBIR framework.	68
6.1	Multi-modality graph construction process	73
6.2	Creating a CAPP graph by pruning edges	75
6.3	Distance and minimum distance features	79
6.4	Relative orientation calculation	80
6.5	Search tree expansion	91
6.6	Search tree expansion with a <i>beam</i> value of 2. \ldots \ldots \ldots	92
7.1	Simulated 2D liver shapes	97
7.2	An example of the OncoPET simulation data	98
7.3	An example PET-CT lung cancer image	100
7.4	Liver shape retrieval (by graph order): mean precision vs. recall $% \mathcal{A}$.	103

7.5	Liver shape retrieval (by graph order): mean precision vs. recall . 10	04
7.6	Example liver shape retrieval	04
7.7	OncoPET retrieval: mean precision vs. recall	05
7.8	Example OncoPET retrieval	06
7.9	PET-CT retrieval: mean precision vs. recall	07
7.10	Example PET-CT retrieval	09
7.11	Graph comparison time by number of tumours	10
7.12	Graph comparison time by beam size	11
7.13	Histograms of graph comparison time	13
7.14	Liver shape retrieval (full data set): precision vs. recall 1	14
7.15	Comparison of complete and CAPP graph structures 1	18
8.1	Retrieval UI design schematic	26
8.2	Retrieval UI implementation	27
	▲	
8.3	Switching between views	29
8.3 8.4	Switching between views	29 30
8.38.48.5	Switching between views. 12 Visualising a graph abstraction. 13 Query refinement. 14	29 30 33
 8.3 8.4 8.5 8.6 	Switching between views. 12 Visualising a graph abstraction. 13 Query refinement. 14 Understanding PET-CT similarity. 14	29 30 33 34
8.38.48.58.6	Switching between views. 12 Visualising a graph abstraction. 13 Query refinement. 14 Understanding PET-CT similarity. 14	29 30 33 34
 8.3 8.4 8.5 8.6 9.1 	Switching between views. 12 Visualising a graph abstraction. 12 Query refinement. 12 Understanding PET-CT similarity. 12 Survey responses – all participants. 14	29 30 33 34 44
 8.3 8.4 8.5 8.6 9.1 9.2 	Switching between views. 12 Visualising a graph abstraction. 12 Query refinement. 12 Understanding PET-CT similarity. 12 Survey responses – all participants. 14 Survey responses – evaluating the baseline system first. 14	29 30 33 34 44 45
 8.3 8.4 8.5 8.6 9.1 9.2 9.3 	Switching between views. 12 Visualising a graph abstraction. 12 Query refinement. 12 Understanding PET-CT similarity. 12 Survey responses – all participants. 14 Survey responses – evaluating the baseline system first. 14 Survey responses – evaluating the proposed system first. 14	29 30 33 34 44 45 45

List of Algorithms

6.1	Graph-Based Query Process
6.2	Modality-Specific Beam-Search Algorithm
6.3	Branch Cost Calculation
8.1	Mapped Vertex ROI Boundary Extraction
A.1	Graph Index Normalisation
A.2	Graph Feature Accumulation
A.3	Feature Collection Initialisation
A.4	Graph Element Feature Accumulation
A.5	Feature Distribution Calculation
A.6	Graph Normalisation
A.7	Vertex and Edge Normalisation

List of Publications

The following publications were produced over the course of the candidature. Most of these publications were based on the work that is presented in this thesis. The other publications were related to projects conducted by my colleagues, to which I made a significant contribution.

Publications Related to Thesis

Published or Accepted

- A. Kumar, J. Kim, W. Cai, S. Eberl, and D. Feng, "A graph-based approach to the retrieval of dual-modality biomedical images using spatial relationships," in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 390–393, 2008.
- A. Kumar, J. Kim, S. Eberl, M. Fulham, and D. Feng, "Content-based retrieval of PET-CT lung cancer data using spatial features," *Journal of Nuclear Medicine*, vol. 50 (Supplement 2), p. 176, 2009.
- A. Kumar, D. Haraguchi, J. Kim, L. Wen, S. Eberl, M. Fulham, D.D. Feng, "A query and visualisation interface for a PET-CT image retrieval system," *International Journal of Computer Assisted Radiology and Surgery*, vol. 6 (Supplement 1), pp. 68–69, 2011.
- 4. J. Kim, A. Kumar, T. W. Cai, and D. D. Feng, "Multi-Modal Content Based Image Retrieval in Healthcare: Current Applications and Future Challenges," in New Technologies for Advancing Healthcare and Clinical Practices, ed. Joseph Tan, pp. 44–59, 2011.
- 5. A. Kumar, J.Kim, L. Bi, , and D. Feng, An image retrieval interface for volumetric multi-modal medical data: application to PET-CT content-based

image retrieval, International Journal of Computer Assisted Radiology and Surgery, vol. 7 (Supplement 1), pp. S475–477, 2012.²

- A. Kumar, J. Kim, M. Fulham, and D. Feng, "Graph-based retrieval of multi-modality medical images: A comparison of representations using simulated images," in *Proceedings of the 25th Symposium on Computer-Based Medical Systems*, pp. 1–6, 2012.
- 7. A. Kumar, J. Kim, L. Wen, and D. Feng, "A graph-based approach to the retrieval of volumetric PET-CT lung images," in *Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5408–5411, 2012.
- A. Kumar, J. Kim, L. Bi, M. Fulham, and D. Feng, "Designing user interfaces to enhance human interpretation of medical CBIR: application to PET-CT images," *International Journal of Computer Assisted Radiology* and Surgery, vol. 8 (Supplement 1), 2013, pp. S72–73.
- A. Kumar, J. Kim, L. Bi, M. Fulham, and D. Feng, "Designing user interfaces to enhance human interpretation of medical content-based image retrieval: application to PET-CT images," *International Journal of Computer Assisted Radiology and Surgery*. EPub May 2013. doi: 10.1007/s11548-013-0896-5
- A. Kumar, J. Kim, W. Cai, D. Feng, "Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data," *Journal of Digital Imaging*. EPub July 2013. doi: 10.1007/s10278-013-9619-2
- A. Kumar, J. Kim, and D. Feng, "Graph-Based Retrieval of PET-CT Images using Vector Space Embedding," in *Proceedings of the 26th Symposium* on Computer-Based Medical Systems. In press.

Under Review

12. A. Kumar, J. Kim, L. Wen, M. Fulham, and D. Feng, "A graph-based approach for the retrieval of multi-modality medical images," submitted to

²This paper received the "CARS 2012 - EuroPACS Poster Award" at the 26th International Computer Assisted Radiology and Surgery Congress and Exhibition.

Medical Image Analysis, under review (undergoing revisions).

Other Publications

Published or Accepted

- J. Kim, A. Kumar, S. Eberl, M. Fulham, and D. Feng, "Interactive point-ofinterest volume rendering visualization of PET-CT data," in *IEEE Nuclear Science Symposium Conference Record*, pp. 4384–4387, 2008.
- 14. D. Haraguchi, J. Kim, A. Kumar, L. Constantinescu, L. Wen, and D. Feng, "A web-based image viewer for multiple PET-CT follow-up studies," in Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5279–5282, 2011.
- J. Kim, A. Kumar, L. Wen, S. Eberl, M. Fulham, D.D. Feng, "Visual tracking of treatment response in PET-CT image sequences," *International Journal of Computer Assisted Radiology and Surgery*, vol. 6 (Supplement 1), pp. 17–18, 2011.
- L. Constantinescu, J. Kim, A. Kumar, D. Haraguchi, L. Wen, and D. Feng, "A patient-centric distribution architecture for medical image sharing," *Health Information Science and Systems*, vol. 1, no. 1, article 3, 2013.

Chapter 1

Introduction

The extraction and visualisation of knowledge from large, ever-expanding repositories is a major challenge in the modern, digital world. This thesis addresses this problem in the domain of medical imaging. In particular, it examines the challenge of finding relevant data from large collections of modern multi-dimensional, multi-modality medical images and answers two key questions: *how can the characteristics of such images be used for search* and *how can the relevant data be interpreted*.

1.1 Motivation

Imaging is a fundamental component of modern medicine. Medical images are used daily for diagnosis [1], treatment planning [2], and assessing a patient's response to treatment [3]. The usefulness of medical imaging has spurred revolutions in the acquisition technologies utilised as part of clinical workflows. A variety of imaging technologies are now available for routine use in hospitals, each with their own benefits for patient management. These revolutionary advances include scanner-based technologies such as x-rays, magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), and single photon emission computed tomography (SPECT); optical imaging technologies such as near infrared spectroscopy; and camera-based technologies like microscopic imaging, and capsule endoscopy. The latest scanners are also capable of acquiring multiple image modalities in a single scanning session; examples, of such devices are the multi-modality PET-CT scanner [4,5] and the newer multimodality PET-MR scanner [6]. The advancement of image acquisition technologies has not been limited to the invention of new devices; modern devices are capable of acquiring images that have a higher resolution than their older counterparts and a larger number of dimensions, e.g., 3D volumetric images and timevarying 4D images.

As a consequence, a vast amount of image data is acquired daily in modern hospitals and the volume of data acquired is expanding at an increasing rate. At the Royal Prince Alfred Hospital in Sydney, the molecular imaging department acquires about 9000 PET-CT images per year, each consisting of several hundred images slices; this is in addition to the acquisition of other imaging modalities. Another often cited example is the number of images stored by the radiology department at University Hospital of Geneva over the past decade [7–9]. Digitisation along with the development of picture archiving and communications systems (PACS) [10] has enabled the storage of these images in ever-expanding digital repositories. Clinical workflows require that imaging data be accessible for use by various clinical staff; PACS provides the capability to share data by exporting images to physical media, such as compact disks, or transferring them across a network. In both cases, the Digital Imaging and Communication in Medicine (DICOM) standard provides the tags necessary for a software system to interpret and render the image as well as meta-data relating to the acquisition [11], such as patient names, scanning times and other identifiers.

These large PACS repositories enable physicians to retrieve indexed images related to a patient, enabling them to consider a patient's historical image data when treating them. Large PACS repositories also provide the opportunity for image-based diagnosis [12], whereby a physician's diagnosis of a patient can be based upon the accumulated knowledge stored in PACS. In addition, PACS can also be used as a resource for the construction of digital teaching files, allowing students and residents to be trained according to collections containing images that are relevant to particular conditions [13]. PACS can also be used for the analysis of data generated by clinical trials, improving productivity and allowing the researchers to "do better science" [14].

Using PACS for these objectives introduces the need to search the repository for images (the targets) that have similar characteristics to a particular image, referred to as the query. In clinical environments, the selection of similar studies is primarily based upon topicality, i.e., whether or not the images contain the same subject matter; judging whether a study is topical is primarily decided upon using the visual characteristics of the images [15]. However, the search capabilities provided by PACS are based on textual keywords, such as patient names, identifiers, and image device. The text descriptions limit the search capabilities of PACS and mean that users must read through clinical reports or already know the keywords of the target images [16,17]. While text-based PACS search is useful when clinical staff already know the identifiers and characteristics of the targets, it is limited for inter-patient comparative studies because it does not consider the visual properties of the images in the repository. Furthermore, the reliance on search via text labels is problematic as even the automatically generated DICOM tags potentially have a high error rate [18].

Labels can be manually assigned to images stored in PACS, such as through using DICOM structured reporting [19]. Labelling is performed by a domain expert, e.g., a radiologist with experience in reading a certain type of image. The expert will first examine the image, potentially in multiple views and dimensions, and then annotate it based on visual characteristics that the expert considers to be important, e.g. location of tumours, anatomical abnormalities, etc. However, this is generally unfeasible for the large volumes acquired in routine clinical practice. The time taken to manually label images is also compounded when modern high resolution, volumetric, and multi-modality images are taken into account. These images contain vast amounts of information that must be examined and then accurately labelled. Economic factors also need to be considered; hiring enough domain experts to label every image would inflate the expenses of hospitals, which are already under budgetary pressures [20]. Manual labelling is also a subjective task with a high dependence on the skill, training, experience, and alertness of the expert labeller [21]. These issues are a severe limitation on the ability of conventional text-based approaches to searching large medical image repositories. A different approach is required, one that is automated and provides a quantitative, non-subjective method of searching medical image repositories.

Content-based image retrieval (CBIR) is an image search technique that complements text-based retrieval through the use of quantifiable image features as search criterion [21]. Features used by CBIR include shape, texture, colour, and the spatial arrangement of objects within an image; these features can be semi-automatically extracted directly from the images, thereby eliminating uneconomical and subjective manual labelling. Unlike classification, CBIR does not use the features to categorise the image into a known group; instead it finds a collection of images that have a similar combination of features.

In the medical domain, CBIR has potential applications in evidence-based diagnosis, physician training, and biomedical research [7] through its ability to find images in repositories that are visually similar to a given query image. Clinical evaluation has demonstrated that the accuracy of a physician's diagnosis can be improved by using a CBIR system to display images similar to the undiagnosed query; the improvement in diagnostic accuracy was largest for less experienced physicians [22]. CBIR has also been shown to have benefits in radiology education [23] by allowing users to retrieve teaching files that consisted of multiple images of a particular clinical case [24]. Furthermore, other investigations have concluded that CBIR could positively impact patient care by providing real-time decision support to physicians [25].

While there have been several advances in medical CBIR of single-modality images, there is a clear gap in the retrieval of multi-modality medical images such as PET-CT and PET-MR [9]. In particular current CBIR research does not utilise complementary information about a patient's state from multiple imaging modalities. Medical CBIR methods for single-modality images have been generally optimised for a particular image modality; the choice of features reflects the information provided by that modality. Directly extending this approach to multi-modality images (by simply extracting optimised features for each modality) does not make use of the greatest asset of such images, the relationships between the constituent modalities. In the case of PET-CT, these relationship features are responsible for improving cancer diagnosis, localisation, and staging, compared to single-modality PET and CT acquired separately [26]. The existence of these complementary relationships provides the opportunity for retrieval based on associations between features of either modality, such as the location of tumours in relation to anatomy. There also exists a gap in the visualisation and presentation of retrieved data for interpretation by physicians [27]. While a few studies have tackled this problem, challenges still remain for modern volumetric and multi-modality images. These images form the most complex challenge for retrieval interpretation because systems must visualise a volumetric data set while enabling users to assimilate the relationships between regions in different modalities (i.e., integrate information from multiple volumes). The increasing clinical utilisation of multi-modality images and the corresponding expansion of medical image repositories means that there is a clear opportunity for advancements in the CBIR of such images.

1.2 Aims and Objectives

The overall aim of the research presented in this thesis is to design a framework for the content-based image retrieval of multi-modality medical images. The framework will exploit the complementary characteristics of the different image modalities to provide image search capabilities on the basis of tumour localisation. Achieving this overall aim will require the fulfilment of the following specific objectives:

- The creation of an image feature representation scheme that indexes features from multiple modalities as well as the relationships between them. The representation scheme will be designed to emphasise tumour localisation in relation to anatomy.
- 2. The derivation of an image similarity measurement algorithm, designed specifically for the new image representation. The algorithm will consider the complementary features in each individual modality.
- 3. The development of a retrieval interpretation technique that enables users to better understand the similarity between the query and the retrieved images.

1.3 Contributions of this Thesis

In this thesis, we propose a novel framework for the CBIR of multi-modality medical images that exploits the complementary characteristics of the different image modalities to introduce new search capabilities in clinical environments. Our framework allows for multi-modality image retrieval on the basis of tumour location relative to anatomy. Our contributions lie in four major areas:

1. Image Representation

We define a novel graph-based representation for multi-modality images, called the Complete Anatomy Proximal Pathology (CAPP) graph, that is able to denote features unique to different modalities while maintaining spatial relationships across modalities. The CAPP graph represents the localisation of disease (from one modality) in relation to anatomy (from another modality) by constraining tumours to spatially related organs using the geometric and topological features used by clinical cancer classification and staging guidelines. It achieves this representation by indexing complementary features from different modalities as well as spatial relationships between regions of interest (ROIs) in different modalities. Our definition of the CAPP graph allows both 2D and 3D multi-modality medical images to be represented.

2. Similarity Measurement

We derive an algorithm for comparing multi-modality images using two criteria: the similarity of image features of elements within the same modality, and the similarity of the disease localisation in each image (i.e., relationships between anatomy and tumours extracted from different modalities). Our algorithm, an adaptation of the graph edit distance, accounts for complementary feature sets for different image modalities and also prevents incorrect mappings between graph elements representing ROIs extracted from different modalities.

3. Retrieval Interpretation

We present a retrieval visualisation and interpretation system for multimodality images by designing a user interface (UI) that was capable of displaying multiple volumetric PET-CT images efficiently through a combination of preprocessing and a rapid load-on-demand approach. Our UI visualises abstractions of clinical PET-CT images to summarise the often complex relationships between tumours and anatomy in each image. We also define interactions for visually-driven image interpretation by exploiting the outputs of our graph comparison algorithm.

4. Multi-Modality Feature Normalisation

We propose a graph feature normalisation scheme to ensure that features with large ranges do not create a bias towards a particular image feature in our retrieval algorithm. The feature normalisation is based upon the distribution of actual feature values in the data set. We normalise features for different modalities separately, ensuring that the normalised value reflects the distribution of a feature within a particular modality.

1.4 Thesis Structure

The remainder of this thesis is structured as follows.

Chapters 2–4 orient the reader by providing the background knowledge necessary to understand the rest of the thesis. Chapter 2 presents an overview of medical imaging and image processing within this domain. Chapter 3 presents an overview of state-of-the-art approaches in CBIR, especially in the medical domain. Chapter 4 provides the reader with a background in graphs, graph-based representations of visual data, and graph similarity calculations.

Chapters 5–9 contain the detailed contributions of the thesis. An overview of our graph based framework is given in Chapter 5. In Chapter 6 we describe the algorithms used to construct CAPP graphs from multi-modality images, the method by which we normalised the image features, and our algorithm for measuring the similarity of multi-modality images, based upon the similarity of their graph structures. The evaluation of our retrieval methodology is presented in Chapter 7. Our retrieval interpretation and visualisation scheme is described in Chapter 8 and evaluated in Chapter 9.

Chapter 10 discusses the capabilities of this research and indicates directions for future investigation. Finally, Chapter 11 summarises the contributions of this thesis.

1.5 Key Terms

The following terms and abbreviations are used throughout this thesis.

Attributed Relational Graph (ARG)

A graph structure where the vertices and edges have been assigned sets of attributes. The attributes for a vertex describe properties of that vertex. Edge attributes describe relationships between the connected vertices.

Axial Plane

An imaginary horizontal plane or cross-section that ranges between the superior (top) and the inferior (bottom) part of the body. It is perpendicular to the coronal and sagittal planes. Also known as the transverse plane.

Computed Tomography (CT)

A medical imaging procedure that computes a greyscale volumetric image using a set of 2D x-rays acquired around a single rotational axis. A CT image is a set of 2D slices or tomograms that form a 3D volume.

Content-Based Image Retrieval (CBIR)

An image search technique where the search is based on image features, such as shape, colour, texture, and the spatial arrangement of objects.

Coronal Plane

An imaginary vertical plane or cross-section that ranges between the anterior (front or ventral) and the posterior (back or dorsal) part of the body. It is perpendicular to the axial and sagittal planes. Also known as the frontal plane.

Digital Imaging and Communication in Medicine (DICOM)

An international standard for managing and transmitting or receiving medical images.

Multi-modality images

Medical images of the same body region acquired with different techniques or modalities. In this thesis, we limit ourselves to multi-modality images acquired by a single hybrid scanner that also co-aligns the two modalities.

PET-CT

A multi-modality medical imaging procedure that sequentially acquires anatomical CT images and functional PET images. Also referred to as PET/CT.

Picture Archiving and Communication System (PACS)

A system that acts as a database for digitised medical images. The images can be accessed over a network using PACS workstations.

Positron Emission Tomography (PET)

A functional medical imaging procedure that constructs greyscale volumetric images by detecting a positron-emitting radiotracer that has been introduced into the subject's body. A PET image is a set of 2D slices or tomograms that form a 3D volume.

Region of Interest (ROI)

A group of pixels in an image that share some similarities or are relevant

as a set, e.g., the collection of pixels that completely encapsulate an object in the image.

Sagittal Plane

An imaginary vertical plane or cross-section that ranges between the left and the right part of the body. It is perpendicular to the axial and coronal planes.

Standard Uptake Value (SUV)

A value for measuring the uptake of PET radiotracers by normalising the original voxel intensities based on PET acquisition parameters, such as dose and time, and subject data, such as mass.

User Interface (UI)

The part of a computer program displayed on a screen that presents users with information and allows them to interact with the program. Most modern UIs are graphical in nature with components such as buttons used for user interaction.

Volume of Interest (VOI)

A 3D region of interest, comprising a group of voxels (3D pixels).

Chapter 2

Medical Imaging and Image Processing

The acquisition of image data is a routine part of modern patient care. A variety of different imaging techniques are used as part of the clinical workflow, with each technique having advantages in different clinical domains and providing insight into different aspects of a patient's condition. This chapter provides an overview of medical imaging and image processing techniques, with a focus on the images and techniques relevant to the research in this thesis.

2.1 Definitions

We begin by defining several terms. An *image* is a collection of picture elements, called *pixels*. A 3D image can also be referred to as a *volumetric image* or *volume*, and comprises a collection of 3D pixels, called *voxels*. Each pixel or voxel also has a value; in greyscale images this value represents the intensity of the pixel, while in colour images this value can be a set of intensities for the red, green, and blue (RGB) channels.

The *pixel resolution* of an image refers to the number of pixels in the image;



Figure 2.1: The pixel and spatial resolution of images: (a) is a 2D image with a pixel resolution of 8×7 and a spatial resolution of 0.5cm $\times 0.5$ cm and (b) is a 3D image with a voxel resolution of $8 \times 7 \times 3$ and a spatial resolution of 1.0mm $\times 2.0$ mm $\times 2.5$ mm.

this can be represented as a single number or by the number of pixels per dimension. For example, the pixel resolution of a 2D image with 2,304,000 pixels is 2.3 megapixels or 1920×1200 pixels (width \times height). The *spatial resolution* of an image refers to size of the detail captured by each pixel. For example, if a volume has a spatial resolution of $1.00 \text{mm} \times 1.00 \text{mm} \times 2.00 \text{mm}$ then each voxel in that volume depicts a region with volume 2.00mm^3 ($1.00 \text{mm} \times 1.00 \text{mm} \times 2.00 \text{mm}$). Images with higher spatial resolutions are capable of depicting finer details. *Contrast resolution* refers to the range of distinct intensities that can be distinguished in an image. A low contrast resolution means that objects with similar but not identical intensities will be difficult to distinguish.

Figure 2.1 depicts an example of a 2D image and a 3D image, envisioned as 2D and 3D arrays or grids, respectively. The 2D image (Figure 2.1(a)) has a pixel resolution of 8×7 and a spatial resolution of 0.5cm $\times 0.5$ cm. The 3D image (Figure 2.1(b)) has a voxel resolution of $8 \times 7 \times 3$ and a spatial resolution of 1.0mm $\times 2.0$ mm $\times 2.5$ mm.

A region of interest (ROI) is collection of pixels that represent an area in an image that holds some importance for a particular application or domain.



Figure 2.2: The region of interest in an image.

Similarly, a volume of interest (VOI) is a collection of voxels representing some important structure. Figure 2.2 depicts an image with two ROIs. Each pixel inside one of the blue shapes is part of that ROI.

2.2 Single Modality Medical Imaging

Single modality medical imaging refers to the "traditional" form of imaging procedures undergone by patients. Every single modality image acquisition produces one type of image, which can potentially be a volume. If multiple types of images are required then different imaging procedures are performed by different scanners during different sessions. Alignment of scans are performed either mentally or, recently, using software to perform image registration (covered in Section 2.4.2). This section will describe several common single modality imaging techniques.

2.2.1 X-ray Imaging

X-ray imaging, also known as radiography, is a medical imaging technology that produces 2D images of the human anatomy. Radiography works by projecting a beam of x-rays through the subject, some of which are absorbed by the body, and detecting the rays that pass through. Since anatomical structures absorb



(a) Axial slice 98(b) Axial slice 139(c) Axial slice 182Figure 2.3: Three CT tomograms (slices) from the same volume.

varying amounts of the radiation, the detector is able to produce a 2D image of the structures in the body. X-rays have seen widespread use in the assessment of fractures [28], the detection of third molars (wisdom teeth) in panoramic dental images [29], and breast cancer screening (mammography) [30]. Modern computed tomography scanners (see Section 2.2.2) also rely upon x-rays for imaging the human body. However, the superimposition of adjacent structures in x-ray imaging and the loss of morphologic image information, such as the 3D arrangement of structures, reduces diagnostic sensitivity and specificity [31].

2.2.2 Computed Tomography

Computed tomography (CT) is a medical procedure that computes a greyscale volumetric image using a set of 2D x-rays acquired around a single rotational axis [32]. A CT image takes the form of a set of 2D slices or tomograms that form a 3D volume. Unlike the traditional x-ray imaging described in Section 2.2.1, the images produced by CT scanners do not superimpose structures on each other. CT scanners are also capable of capturing images with very high spatial resolutions, potentially less than 1mm per dimension.

Three axial (top-to-bottom) slices from a single CT volume are shown in Figure 2.3. The differences in voxel intensities can be clearly seen by the high intensity grey values of the bones in Figure 2.3(a), the low intensity of the air within the lungs in Figure 2.3(b), and the intensity of the soft tissues (liver,
spleen, etc.) in Figure 2.3(c).

CT images are primarily used for the detection and analysis of anatomical conditions, such as airway analysis in patients suffering from obstructive sleep apnea [33] or assessing emphysema [34]. They are also used for measuring tumour growth [35] and determining the growth rates of lung nodules [36]. However, benign or malignant tumours cannot be easily differentiated on the basis of CT images alone and a different procedure, such as a biopsy or positron emission tomography (see Section 2.2.3), is usually required.

2.2.3 Positron Emission Tomography

Positron emission tomography (PET) is a functional medical imaging procedure that constructs greyscale volumetric images by detecting a positron-emitting radiotracer that has been introduced into the subject's body [37]. In this thesis, we only consider PET images that use the commonly utilised radiotracer ¹⁸F-Fluorodeoxyglucose (FDG).

Voxel intensity in an FDG-PET image indicates the glucose metabolism at that location in the body; this is useful for characterising the nature of lesions because malignant tumours appear as regions with abnormally high intensities [38]. For this reason FDG-PET images have a diagnostic and prognostic accuracy between 80-90%, making them better at detecting malignant cancers compared to anatomical imaging modalities like CT [39]. However, these images have a low spatial resolution and are unable to capture fine details and have a lower signalto-noise ratio than modalities like CT [40]. There is also insufficient anatomical information in PET for the accurate localisation of lesions [39].

Furthermore, the voxel values in PET images do not naturally correspond to a physical characteristic (unlike CT voxel values, which are related to the x-ray absorption of different materials). This makes the comparison of PET values across studies unreliable, even if the scans are of the same patient [41].



Figure 2.4: Slices from different planes of the same PET volume.

The standard uptake value (SUV) is a value for measuring the uptake of PET radiotracers by normalising the original voxel intensities based on PET acquisition

parameters, such as dose and time, and subject data, such as mass or weight.

Figure 2.4 shows two slices from an FDG-PET scan. Figure 2.4(a) is a slice from the volume when viewed in the coronal (front-to-back) plane. The two regions of high intensity indicated by the blue arrows are tumours. All other high intensity regions correspond to areas with natural high glucose metabolism, e.g. the brain. Figure 2.4(b) shows the image in the sagittal (side-to-side) plane. The tumours are not visible from the position from which this slice was taken. These views were selected to show the variation of PET voxel intensity across the body.

2.2.4 Magnetic Resonance Imaging

Magnetic resonance (MR) imaging is a procedure that constructs an image by detecting the atomic nuclei in the body that have rotating magnetic fields due to interactions with the powerful magnetic field produced by the scanner [42]. MR imaging is non-ionising. In addition, MR images have a better contrast resolution



(a) Axial view.(b) Coronal view.Figure 2.5: T2 weighted brain MR image slices.

than CT images and are therefore able to distinguish tissues that are similar but not identical. Different tissue types can be emphasised by varying acquisition parameters, e.g., the pulse sequences, to obtain MR images with different image contrasts (e.g. T1 or T2 weighted scans etc.). Other variants include diffusion MR, which measures the diffusion of water molecules in various tissues, as well as functional MRI, which captures dynamic neural activity.

Figure 2.5 shows two T2 weighted MR image slices of the brain. The contrast resolution of the images is visually quite clear from the different levels of grey in both the axial slice (Figure 2.5(a)) and the coronal slice (Figure 2.5(b)).

2.3 Multi-Modality Medical Imaging

Multi-modality imaging refers to the acquisition of different types (modalities) of images of the same body region. While this sort of imaging can be provided by imaging acquisitions from multiple single-modality imaging devices, in this thesis we only consider multi-modality images acquired by a single scanner during a single session. These images are hardware co-aligned by the scanner. This subsection describes the multi-modality medical imaging modalities relevant to this thesis.



(a) Greyscale PET (b) Greyscale CT (c) Fused PET-CT Figure 2.6: A PET-CT image.

2.3.1 PET-CT

PET-CT is the sequential acquisition of CT and PET volumes in the same scanner during the same imaging session [4,5]. The two volumes acquired by the scanner have different pixel, spatial, and contrast resolutions. Figure 2.6 shows the axial images acquired from a combined PET-CT scanner. Figures 2.6(a) and 2.6(b) are the PET and CT images, respectively. Figure 2.6(c) depicts the fusion of these images after scanner parameters have been used to transform them into the same coordinate space.

Combined PET-CT scanners offer considerable advantages over their single modality counterparts. The total image acquisition time is significantly shorter leading to better instrument utilisation and a higher patient throughput [26]. Furthermore, studies have shown that PET-CT is more sensitive than either PET or CT performed alone, and that the CT scan adds sensitivity to the PET image [43]. Overall, PET-CT provides improved tumour diagnosis, localisation, and staging, compared to single modality PET or CT [26,43]. The clinical usefulness of PET-CT and the trends in PET adoption indicate that in the very near future all PET studies will be in the form of PET-CT images [39,44].

The value of PET-CT arises from its ability to present complementary anatomical (CT) and functional (PET) information. The spatial co-alignment of the two modalities performed by the scanner enables a physician to see the relationship between the anatomical and functional information, such as whether a lung tumour is invading an adjacent structure. An essential challenge of PET-CT image processing research is capitalising on these complementary features and spatial relationships.

2.3.2 PET-MR

PET-MR is the simultaneous acquisition of MR and PET volumes in the same scanner, during one imaging session, and without sacrificing the image quality of either modality [6]. MR does not use ionising radiation meaning that PET-MR can be used in situations where radiation exposure is a concern, such as serial studies. Furthermore, MR images have a higher contrast resolution when compared to CT, thereby enabling even better soft tissue definition than was possible with PET-CT [45]. PET-MR has great potential for brain research and for assessing and translating new treatments into clinical application [46].

2.4 Image Processing

This subsection describes several categories of image processing algorithms. We focus on fully or semi-automatic processes, even when completely manual approaches are possible. Our attention will be on those processes that are necessary for understanding the later chapters of this thesis.

2.4.1 Segmentation

Segmentation is the process by which the pixels or voxels of an image are partitioned into different regions that share specific characteristics [47]. It is generally used to identify boundaries, regions, or individual objects within an image. Other



Figure 2.7: Segmenting a chest CT image using [48] with smoothing.

image processing algorithms can then be applied to the different segmented ROIs individually instead of being applied to the entire image.

Pixel analysis is a core component of image segmentation. Common segmentation approaches include: pixel thresholding [48, 49], region growing [50], edge detection [51], fuzzy clustering [52], graph cuts [53], and statistical shape modelling [54]. The segmentation algorithms used in this thesis are mainly thresholding and region growing. Thresholding separates pixels into groups that have similar characteristics, such as colours or intensities. Region growing analyses the neighbouring pixels of one or more seed points (with the initial points usually selected manually) and recursively expands a region while the neighbouring pixels share common properties.

Figure 2.7 depicts an example of image segmentation. A well-established iterative thresholding algorithm [48] combined with smoothing has been applied to the original chest CT image (Figure 2.7(a)) to segment the two lungs (Figure 2.7(b)) from the surrounding soft tissue, mediastinum, ribs, and other anatomical structures.

2.4.2 Registration

Registration is the process of transforming two images of the same scene or object into the same coordinate space [55]. It is used to integrate information from images acquired from different devices, from different viewpoints, or at different



(a) Source image. (b) Target image. (c) Registered image.

Figure 2.8: Registering two CT images. The source image (a) has been transformed into the coordinate space of the target image (b) using non-rigid b-spline registration [57], producing the transformed image (c).

times. A transformation matrix that maps pixels between the images is generated during the procedure. Registration is an optimisation process that attempts to maximise the similarity of corresponding parts of the two images while minimising the degree of transformation applied.

Registration algorithms fall into several categories. Rigid or linear registration algorithms apply global transformations to align pixels in the entire source image to the target image, while non-rigid algorithms are elastic and allow independent local morphing of different parts of the image [56]. They are useful for medical image analysis where it is important to register different structures independently between images.

Figure 2.8 depicts two images registered using non-rigid b-spline registration (implemented in the Elastix toolbox [57]). The source image (Figure 2.8(a)) has been transformed to the coordinate space of the target image (Figure 2.8(b)); the result is depicted in Figure 2.8(c).

2.4.3 Visualisation

Visualisation refers to the process by which images are displayed. The simplest form of visualisation is rendering an image 'as is', without any transformations to emphasise any particular aspect. More complex visualisations create a transformed version of the original image data to highlight a certain component of the image or to meet technical limitations. One example of the former is applying colour look up tables (LUTs) to greyscale images enabling different parts of the images to be distinguished by colour as opposed to grey-level intensity. An example of the latter is capturing a certain point of view in a 3D scene to enable its display on a computer monitor, a 2D surface.

There are several common medical imaging visualisation techniques. Transfer functions [58] create a mapping between voxel intensities, colour, and opacity, thereby applying a level of transparency to voxels of a particular intensity, allowing structures inside the volume to be rendered instead of just the surface of the volume. Direct volume rendering [59] is a method that projects a 3D volume into 2D view given a specific viewpoint, by considering the voxel intensities, colours, and opacities as specified by transfer functions. A specific form of volume rendering, maximum intensity projection (MIP), displays the voxels in the image that have the highest intensities from a particular viewpoint [60] but sacrificing depth information in the process.

Chapter 3

Content-Based Image Retrieval

This chapter provides an overview of content-based image retrieval. We present a survey of the current state-of-the-art, with a focus on its application to the medical domain, and identify the gaps in existing technologies.

3.1 Measuring Image Similarity

What Renaissance paintings are similar to da Vinci's *Mona Lisa*? Raphael's *Portrait of a Young Woman with a Unicorn* is certainly one possibility. As seen in Figure 3.1, the paintings share a number of similarities: they are both portraits of women, the background in each painting features a landscape with mountains, and each of the women is dressed in Renaissance fashion. Yet to the human eye, the images are undeniably not the same: the *Mona Lisa* has darker colours, the landscape features a winding road, the subject is not wearing a pendant, and she is not holding a unicorn! If similarity was based on colour alone, then Raphael's *Portrait of a Man* could be considered similar to the *Mona Lisa*. However, other elements of the painting are quite different: much of the background landscape is not visible, the subject is facing directly forward, and, perhaps most obvious of all, the subject is male!



(a) The Mona Lisa



(b) Portrait of a Young Woman with a Unicorn





(c) Portrait of a Man

What if the question asked us to rank Raphael's paintings by their similarity to the *Mona Lisa*? An objective answer to this question could be obtained if every element of these images could be quantitatively measured and compared. In this case the old adage "a picture is worth a thousand words" holds true; the amount of information encoded by images is so high that it is impractical to describe it all manually. The problem of measuring the information in images is magnified if the scope of our earlier question is increased to all Renaissance art, let alone all art produced in the history of mankind.

3.2 Content-Based Image Retrieval

Content-based image retrieval (CBIR) is an image search technique that does not rely upon the use of manually assigned annotations. Instead, CBIR complements text-based retrieval through the use of quantifiable and objective image features as the search criterion. The features used by CBIR include, but are not limited to, colour, texture, shape, and the spatial arrangement of ROIs within the images [21]. These features can be automatically or semi-automatically extracted directly from the images, thereby eliminating uneconomical and subjective manual labelling. Essentially, CBIR measures the similarity of two images based on the similarity of the properties of their visual components, e.g. similarity based on colour distribution. CBIR's non-reliance on labels has made it ideal for large repositories where it is not feasible to manually assign keywords and other annotations. The objective features used by CBIR mean that it is also possible to show not only *what* images are similar but also to explain *why* they are considered to be similar in a non-subjective manner. The *what* is essentially the set of retrieved images and is provided by every CBIR algorithm.

The major challenges for CBIR include the application-specific definition of similarity (based on users' criterion), extraction of image features that are relevant to this definition of similarity, and organising these features into indices for fast retrieval from large-scale repositories [21,61–63]. The choice of features becomes a critical task when designing a CBIR system because it is closely related to the definition of similarity. Features fall into several categories. General-purpose features can be extracted from almost all images but are not necessarily appropriate for all applications, e.g., colour is inappropriate for greyscale images. Applicationspecific features are tuned to a particular problem and describe characteristics unique to a particular problem domain; they are semantic features intended to encode a specific meaning [21]. Global features capture the overall characteristics of an image but fail to identify important visual characteristics if these characteristics occur in only a relatively small part of an image. Local features describe the characteristics of a small set of pixels (possibly even one pixel), i.e. they represent the details. In recent years, there has been a shift towards the use of local features largely driven by the belief that most images are too complex to be described in a general manner; however, the combination of local and global features remains an area of investigation for practical computer vision applications [63].

An underlying assumption of most CBIR systems is that the chosen image features used are sufficient to describe the image accurately. The choice of image features must therefore be made to minimise two major limitations: the *sensory gap* and the *semantic gap* [21]. The sensory gap is the difference between the object in the world and the features derived from the image. It arises when an image is noisy, has low illumination, or includes objects that are partially occluded by other objects. The sensory gap is further compounded when 2D images of physical 3D objects are considered; some information is lost as the choice of viewpoint means an object may occlude part of itself. The semantic gap is the conflict between the intent of the user and the images retrieved by the algorithm. It occurs because CBIR systems are unable to interpret images; they do not understand the 'meaning' in the images in the same way that a human does, i.e., CBIR is performed on the basis of image features not image interpretations.

The large volume of modern image repositories and high feature dimensionality of images has also contributed to challenges in efficient real-time retrieval. In many cases, it is no longer viable to compare a query to every element of the data set. Efficient indexing schemes are necessary to store and partition the data set so it can be accessed and traversed quickly, without needing to visit or process irrelevant data; alternatively, the search space can be pruned by using only a subset of the features or applying weights to features [63]. The large data volumes also mean that exact search paradigms, which look for images in the data set that exactly satisfy *all* query criteria, may no longer be viable. This has led to the rise of approximate search schemes, which rank the images in the data set according to how well they satisfy the search criterion [21]. Perhaps the most well-known approximate scheme is k-nearest-neighbour search, which retrieves the k most similar (highly ranked) images as measured by distance from the query in the feature space. It is possible that some images retrieved by approximate search paradigms will fail to meet the expectations of the users. Precision and recall are two quality measures defined to calculate the accuracy of an approximate search paradigm. Precision refers to the proportion of retrieved images that are relevant, i.e., the proportion of all retrieved images that the user was expecting. Recall is the proportion of all relevant images that were retrieved, i.e., the proportion of similar images in the data set that were actually retrieved. The ideal case would be a retrieval system that achieves both 100% precision and recall. The reality is that most existing algorithms fail to find all similar images, and many of the retrieved images contain dissimilar images or false positives.

Early examples of CBIR use include IBM's Query By Image Content (QBIC)¹ system [64], used to search for famous artworks, as well as the Virage framework [65] and Photobook [66]. More recently, Google Search by Image² used the points, colours, lines, and textures in images uploaded by users to find similar images [67]. This recent development means that CBIR is a technology that is available to the masses.

In recent years, a paradigm shift has changed the focus of CBIR research towards application-oriented, domain-specific technologies that would have greater impact on daily life [63]. Due to advancements in acquisition technologies, ongoing CBIR research has moved towards images with more dimensions, with an aim towards increasing image understanding. Modern medical imaging is one such domain, where the retrieval of multidimensional and multi-modality images from repositories of diverse data has potential applications in diagnosis, training, and research [7]. The contents of medical images exhibit complex characteristics: there is a high variability in the detail of anatomical structures across patients, misalignment of structures can occur in volumetric and multi-modality images,

¹http://wwwqbic.almaden.ibm.com/

²Click the camera icon in the search bar on http://images.google.com/.

some imaging modalities suffer from low signal-to-noise ratios, and occlusion of structures is a common occurrence. In addition, there can be large variability even among patients with the same health condition [68]. It is essential that the characteristics of particular medical images be taken into account when designing CBIR systems for them.

Most CBIR frameworks, even those for medical images, follow similar sequence of processes. Visual features are extracted from the images and indexed for searching. A similarity measurement algorithm is defined to compare a query with the indexed images. The measurements can then be used to rank the images in order of similarity, or can be used to classify the images as 'similar' or 'not similar'. This ranking is then displayed to the user. In many cases, the integration of user feedback allows further refinement of the results to overcome the semantic gap. The following surveys provide detailed overviews of general CBIR frameworks and components: [21, 61–63]. We provide a brief overview of research in these processes in the following subsections.

3.2.1 Features and Representations

Visual feature extraction is the process by which an image is analysed and the properties of its contents are measured, and forms the foundation upon which CBIR stands [62]. The features used for CBIR are generally dependent upon the specific domain and for a particular aim [17]. These features can either be global (calculated from the entire image) or local (calculated from specific ROIs). Image segmentation (see Section 2.4.1) is generally used to define the ROIs for local feature calculation. The representation of a particular image is closely related to the features that need to be represented.

The use of colour features have been attributed to the three-dimensional domain it offers compared to the single dimension of grey-level images [21]. Colour features are relatively robust to background complications, and are independent to image size and orientation [62]. Colour information can be represented by colour histograms [69], colour moments [70], and several other approaches (see [21,62,71]).

Texture is an innate property of almost all surfaces, containing information about the structural arrangement of these surfaces and their relationships to other surfaces. The most common and widely used texture features are the Haralick texture features [72] extracted from a pixel co-occurrence matrix. While originally intended for 2D images, 3D Haralick features can be extracted by calculating the co-occurrence matrix in the thirteen unique orientations [73]. Texture can also be extracted from the coefficients of wavelet transforms [74, 75].

Shape features capture the geometric details within the image [62]. The seven Hu moments [76] are invariant for transformations and as such are ideal for situations where shapes may be rotated, translated, or have varying scales in different images. Other methods describe shapes in reference to the boundary, e.g., shock graphs [77] represent the perturbation of the shape boundary. Two approaches for 3D object retrieval were proposed in [78] based on object surface curvature, and correlograms of the objects from different viewpoints. The study also proposed a decomposition of 3D objects into a set of components, enabling 3D object retrieval based on individual or sets of parts. A review of 3D shape descriptors used for retrieval can be found in [79].

The structure or layout of an image can be represented by the relationships between entities, often in the form of graphs, trees, or hierarchies. These structures can also index other features (colour, texture, etc.) while enabling the representation of relationships such as spatial arrangements [80], or hierarchical ordering [81, 82]. When objects are adjacent, the intensity profile of the local neighbourhood of tumours can also provide relationship features [83]. Other methods for representing spatial relationships include the use of triangular spatial relationships (angles between groups of entities) [84], matrices indicating the relative cardinal or ordinal (compass) directions of objects [85], and complex strings for detailing the topological and geometric interactions between objects [86].

3.2.2 Similarity Measurement

The similarity of two images is essentially a decision problem that interprets the differences between the feature sets of individual images [21]. In many cases, the similarity functions, and optimisations such as weights, depend on the domain for which the CBIR framework is being designed. The result of similarity measurement can either be a ranking of images based upon the degree of correspondence with the query image, or a binary classification (similar or dissimilar to the query).

When the features are in a vector space, such as in the case of the Haralick texture features, the measurement can occur from the distance function D_p as follows:

$$D_p\left(\mathbf{Q},\mathbf{S}\right) = \left[\sum_{i=1}^N \left(q_i - s_i\right)^p\right]^{\frac{1}{p}}$$
(3.1)

where \mathbf{Q} and \mathbf{S} are a query and data set feature vector, respectively; N is the length of the vectors \mathbf{Q} and \mathbf{S} ; q_i and s_i are *i*-th feature values in the vectors \mathbf{Q} and \mathbf{S} , respectively; and, p is the order of the equation. D_1 and D_2 are the Manhattan (city-block) and Euclidean distances, respectively. When individual features are given weights, this function changes to:

$$D_p\left(\mathbf{Q}, \mathbf{S}, \mathbf{W}\right) = \left[\sum_{i=1}^N w_i \left(q_i - s_i\right)^p\right]^{\frac{1}{p}}$$
(3.2)

where \mathbf{W} is a vector of weights, and w_i is the *i*-th element of \mathbf{W} (and thus the weight of the *i*-th feature in \mathbf{Q} and \mathbf{S}).

There is a danger that features with large ranges or values can have a greater

impact on similarity measurement, i.e., if $x \gg y > 0$ then a feature with the range [0, x] can potentially have a greater impact on the similarity measure than a feature with the range [0, y]. Normalisation is a form of weighting that equalises the contribution of each feature to the similarity function, resulting in a measure that is not inherently biased towards a particular feature. Numerous feature normalisation schemes for image retrieval applications are presented in [87].

Structural features are measured by comparing the differences in the organisation of the relationships between ROIs. Essentially such similarity methods attempt to create a mapping between elements of two structures, e.g., ROI A in image I corresponds to ROI B in image J. Relaxation labelling [88] iteratively assigns labels to structures based on a set of probabilities derived from contextual constraints. Graph isomorphism methods [89,90] attempt to find a correspondence between elements in the query and data set graphs (see Section 4.3 for further details).

A similarity measure can be trained to favour particular interpretations. Neural networks, Bayesian classifiers, support vector machines (SVMs), and Hidden Markov Models can be used for this purpose [7]. Further information on various similarity measurements can be found in [7,21,63,91].

3.2.3 Display and Feedback

The semantic gap necessitates an active user as part of the image retrieval process. A human must be presented with the retrieved images so that a semantic interpretation can be performed, i.e., the user must confirm whether the results of the retrieval matched his or her intent. A retrieval system must therefore provide some way for the user to meaningfully view and interact with the displayed data.

Most CBIR systems display the retrieved images in a grid sorted from most similar to least similar. The user then must browse through these retrieved images to locate the most similar images, according to their semantic interpretation. It was shown that arranging images based upon their similarity assisted in picture selection tasks [92]. An overview of general exploratory search systems is given in [93] and a survey of browsing models can be found in [94].

There have been several recommendations about the functionality of search interfaces [95] and the consideration of human factors when displaying information [96]. These recommendations include details about the presentation of both query and retrieved data, the use of multiple views to provide a comprehensive perspective of the retrieved data, and the use of abstractions to provide an overview of complex data. The recommendations also state that the user should always have control over the presentation and ordering of the retrieved information. The properties of elements, the relationships between them, and supplementary data should be displayed to enable user semantic interpretation. The ability to iteratively refine queries to narrow down the search space is also valuable.

One of the recommendations refers to the ability to refine the query, either through filtering, sorting or iterative searching. Relevance feedback [63] is a mechanism by which the user is able to iteratively improve the pertinence of the retrieved images by marking retrieved images as 'relevant' or 'not relevant'. This enables the user to narrow the retrieval to those most relevant to their semantic interpretation of the query [91]; it is a means to bridge the semantic gap. While relevance feedback can be implemented in a number of ways (see [97] for a review), there must be a focus on displaying the images and features transparently to the user, i.e., showing the user which features have made an image similar [21].

3.3 Medical Content-Based Image Retrieval

PACS and other hospital information systems (HIS) store a large variety of information, ranging from clinical measurements (age, weight, blood pressure) to free text reports, test results, and images. These systems have been designed to support higher quality patient care through the effective and efficient management of patient data by enabling physicians to access relevant data in a timely manner [98]. Diagnostic decision making has traditionally involved using evidence from a patient's data coupled with the physician's prior experiences with similar cases [99]. The introduction of PACS has created an opportunity to leverage large repositories to support clinical decision making, by analysing the stored data for similarities in diseases-specific information across patients [9]. This is especially useful for non-specialists that may have had very limited experience with particular cases.

CBIR technologies are seen as promising methods for utilising this diverse and information rich medical imaging repositories for these purposes. Several studies have already demonstrated the potential benefits of CBIR in clinical applications. Clinical evaluation of the ASSERT CBIR system for high resolution CT lung images [100] showed an improvement in the accuracy of the diagnosis made by physicians [22]. Another study for liver CT concluded that CBIR could be used to provide real-time decision support [25]. CBIR was also shown to have benefits when used as part of a radiology teaching system [23].

In the following subsections, we review CBIR developments that have enabled medical image access for clinical applications. Existing reviews [7, 8, 101, 102] mainly considered the differences in features and algorithms applied to medical imaging, and the domains in which they were applied. We take a different approach by describing the evolution of CBIR methods for the retrieval of modern multi-dimensional and multi-modality medical images. In particular, we survey different applications of and approaches to medical CBIR in five main categories: 2D image retrieval, retrieval of images with 3 or more dimensions, using non-image data to enhance the retrieval, retrieval from diverse data sets, and multi-image (and multi-modality) retrieval. We use these categories as a framework for discussing the state-of-the-art, focusing on the characteristics and modalities of the

Type of Data	Studies
2D Images	radiographs: $[103-105]$; spine x-rays: $[106-112]$; cervico- graphs: $[113]$; mammograms: $[114-116]$, $[117, 118]^{a}$; retino- pathy: $[116]$, $[117, 118]^{a}$
3D+ Images	CT: [22,100,119], [25] ^a ; MRI: [82,120,121]; dynamic PET: [122,123] ^a ; PET-CT: [124–130], [131–140] ^c
Non-Image Data	text: $[122, 123, 141-147]^{b}$, $[148, 149]$; annotation or ontology: $[25, 150, 151]^{b}$; others: $[117, 118]^{b}$
Multiple Images	ImageCLEF: [152–156]; pathology: [157]; general [158, 159]; PET-CT: [124–130], $[131-140]^{c}$

Table 3.1: Medical CBIR Studies Divided by Data Types

^a Also used non-image data.

^b Also used image data.

^c Based on work described in this thesis.

information used during medical image retrieval. Table 3.1 provides a brief summary of the studies that we will examine in this review and the types of data used during retrieval.

3.3.1 2D Medical Image Retrieval

The majority of CBIR research on 2D medical images has focused on radiographic images, such as x-rays and mammograms. The Image Retrieval in Medical Applications (IRMA) project³ has been a sustained effort in the CBIR of radiologic images for medical diagnosis systems. The IRMA approach is divided into seven interdependent steps [103]: (i) categorisation based on global features, (ii) registration using geometry and contrast, (iii) local feature extraction, (iv) category and query dependent feature selection, (v) multi scale indexing, (vi) identification of semantic knowledge, and (vii) retrieval on the basis of the previous steps. The IRMA method classifies images into anatomical areas, modalities and viewpoints and provides a generic framework [104] that allows the derivation of flexible implementations that are optimised for specific applications.

³IRMA Homepage (English): http://www.irma-project.org/index_en.php

Other approaches for radiograph retrieval have tried to group features into semantically meaningful patterns. In one such study [105], multi-scale statistical features were extracted from images by a 2D discrete wavelet transform. These features were then clustered into small patterns; images were represented as complex patterns consisting of sets of these smaller patterns. Experimental results revealed that the method had significantly higher precision and recall compared to two conventional approaches: local and global grey-level histograms.

A series of studies [106–112] investigated every component of CBIR for spine x-ray retrieval, including feature extraction [107, 108, 111], indexing [112], similarity measurement [109, 112], and visualisation and refinement [110]. The initial methods of matching whole vertebrae shapes [107, 108] had a major drawback: in 2D x-rays, regions of the vertebrae that were not of pathologic interest could obscure differences between critical regions. Partial shape matching [109] was proposed as a way to deal with occlusion when comparing incomplete or distorted shapes. An application-specific feature, the 9-point landmark model used by radiologists and bone morphometrists in marking pathologies, was localised to improve the computational performance of their algorithm for partial shape matching. In experiments their method achieved a precision greater than 85%. While the users could apply weights to angles, lengths, and the cost to merge points on the model, it was difficult to determine the effect of these weights on the retrieval results.

This was resolved in a later study [110], where a web-based spine x-ray retrieval system contained a query editor that allowed a user to alter the appearance of a shape and to assign weights to points on the shape to emphasise their importance. The integration of relevance feedback further improved the performance of the algorithm. Originally 68% of the retrieved images were relevant (what the user expected); three iterations of feedback increased this by a further 22%. Assigning weights to parts of the shape allowed the user to specify *why* the images were similar. The web-based shape retrieval algorithm was demonstrated to also work with uterine cervix images; the system was able to distinguish between three tissue types with an accuracy of 64% [113].

The spine retrieval framework was further enhanced with the introduction of several domain-specific features: the geometric and spatial relationships between adjacent vertebrae [111]. Combining these features with a voting consensus algorithm improved retrieval accuracy by about 8%. To improve the speed of the retrieval, Qian et al. [112] indexed the images by embedding the shapes in a Euclidean space. This index resulted in significantly faster retrieval times (0.29s compared to 319.42s). It was also discovered that the embedded Euclidean distance measure was a very good approximation of Procrustes distance used previously: the first 5 retrieved images were identical for both methods over 100 queries.

Korn et al. [114] proposed a tumour shape retrieval algorithm for mammography images. In particular, the study introduced application-specific features to model the 'jaggedness' of the periphery of tumours; tumours were represented by a pattern spectrum consisting of shape characteristics with high discriminatory power, such as shape smoothness and area in different scales. This was done to differentiate benign and malignant masses, which are more likely to have higher fractal dimensions. Experiments on a simulated data set revealed that the proposed application-specific approach achieved 80% precision at 100% recall. Their use of pruning to reduce the search space resulted in computational performance that was up to 27 times better than sequential scans of the entire data set.

In [115] a boosting framework was used to learn a distance metric that preserved both semantic and visual similarity during medical image retrieval. Initially, sets of binary features for data representation were learned from a labelled training set. To preserve visual similarity, sets of visual pairs (pairs of similar images) were used alongside the binary features for training the distance function. The proposed approach had a higher retrieval accuracy than other retrieval methods on mammograms and comparable accuracy to the best approach on the x-ray images from the medical data set of the Cross Language Evaluation Forum's imaging track (ImageCLEF)⁴. The retrieval framework performed more consistently than other state-of-the-art approaches across different data sets due to its ability to learn feature sets and distance functions optimised for a particular data set.

3.3.2 3D+ Medical Image Retrieval

In recent years, many 2D retrieval algorithms have been adapted for use in 3D medical image retrieval. Perhaps the most well-known example is the ASSERT system [100], which retrieved a volumetric high resolution CT (HRCT) image on the basis of *key slices* selected from the volume. This essentially reduced a 3D image retrieval problem to 2D retrieval. The system retrieved images with the same type of lung pathology (e.g. emphysema, cysts, metastatic classification etc.), preferably within the same lung lobe as the query. During the query process, a physician would mark a pathology bearing region in a HRCT lung slice; grey-level texture features, as well as other statistics, were then extracted from these regions. Relational information about the lung lobes was also captured. In experiments, the ASSERT system achieved a retrieval precision of 76.3% when matching the type of disease; this dropped to 47.3% when pathology location was also considered. During clinical evaluation [22], physicians used the ASSERT system to retrieve and display four diagnosed cases that were similar to an unknown case; this was shown to improve the accuracy of their diagnosis.

An improvement to the ASSERT system involved a two-stage unsupervised feature selection method to "customise" the query [119]. During the first stage, the features that best discriminated different classes of images were used to classify the query into the most appropriate pathology class. In the second stage,

⁴ImageCLEF Homepage: http://www.imageclef.org/

the features that best discriminated between images within a class were used to identify the "subclass" of the query, i.e., to find the most similar images within the class. The customised query approach had an effective retrieval precision of 73.2% compared to 38.9% using a single vector of all the features. The study demonstrated that finding images on the basis of class was not enough; there was a need to also find the most similar images within a particular class.

Local structure information in ROIs was used for the retrieval of brain MR slices [120]. Two feature sets for the representation of structural information were compared. The first, local binary patterns (LBPs), treated every local ROI equally. The other, Kanade-Lucas-Tomasi (KLT) feature points, gave greater emphasis to the more salient regions. The results revealed interesting insights about the tradeoffs inherent in structure-based retrieval. LBPs were very dominant when spatial information was included, and its accuracy was consistently higher than its rivals in experiments involving pathologies or other anomalies. The experiments also showed that accuracy was degraded when KLT points were not matched.

Petrakis [121] proposed a graph-based methodology for retrieving MR images. Each image was represented by an attributed graph; vertices represented ROIs while edges represented relationships between ROI. Their results showed that a similarity measure based on the concept of graph edit distance achieved the best retrieval precision, at the cost of computational efficiency. Alajlan et al. [82] proposed a tree representation that achieved improved computational performance by only indexing relationships between ROIs that were included (completely surrounded) within other ROI.

Dynamic PET images consist of a sequence of PET image frames acquired over time. Cai et al. [122] proposed a CBIR system that utilised the temporal features in these images. They exploited the activity of pixels or voxels across different time frames by basing their retrieval on the similarity of tissue time activity curves (TTACs) [160]. In [122], three query input methods were allowed: textual attributes, definition of a query TTAC, and a combination of these features. Kim et al. [123] extended this retrieval to 4 dimensions (3 spatial and 1 temporal) by registering 3D brain images to an anatomical atlas, and defining the structures to search using the atlas' labels.

3.3.3 Retrieval Enhancement Using Non-Image Data

Text information is a common complement to image features in general CBIR research [161] as well as medical CBIR research. Several examples of studies including non-image data have already been described [122,123]. Textual information has also been used to complement several studies that were part of the ImageCLEF medical challenge or used the same data [141–147].

An initial approach to using text as the input query mechanism for image data together was presented by Chu et al. [148]. The spatial properties of ROIs and the relationships between them were indexed in a conceptual model consisting of two layers. The first layer abstracted individual objects from images, while the second layer modelled hierarchical, spatial, temporal, and evolutionary relations. The relationships represented the users' conceptual and semantic understanding of organs and diseases. Users constructed text queries using an SQL-like language. Each query specified ROI properties, e.g., organ size, as well as relationships between ROIs. This retrieval approach was expanded in [149] with the introduction of a visual method for query construction and by the inclusion of a hierarchy for grouping related image features.

Rahman et al. [146] presented a technique that used the correlation between text and visual components to expand the query. Their comparison of text, visual, and combined approaches revealed that the text retrieval had a higher mean average precision than the purely visual method, while the combined method outperformed both text and visual features alone. This outcome was also visible in a comparison of different retrieval algorithms in [147] but could be explained by the nature of the data set that was used. The medical images in the ImageCLEF data set were highly annotated and this made text-based retrieval inherently easier than purely visual approaches.

A comparison of text, images, and combined text and image features was conducted in [150], using a data set that was not as well annotated. The text features were extracted from the caption of the images in the document, as well as paragraphs referring to those images. The experiments consisted of an indexing task that produced a single IRMA annotation for an image and a retrieval task that matched images to a query. The results showed that image analysis was better than text for both indexing and retrieval, though there were a few circumstances where indexing performed better with text data. The results also revealed that caption text provided more suitable information than the paragraph text. While combined image and text data seemed beneficial for indexing, the retrieval accuracy was not significantly higher than that of using images alone.

A preliminary clinical study [25] evaluated different features for the retrieval of liver lesions in CT images. In particular, the study compared texture, boundary features, and semantic descriptors. Twenty-six unique descriptors, from a set of 161 terms from the Radlex terminology [151], were manually assigned by trained radiologists to the 30 lesions in the data set; each lesion was given between 8 to 11 descriptors. The semantic descriptors were a feature that explained *why* images were clinically similar. The similarity of a pair of lesions was defined as the inverse of a weighted sum of differences of their respective feature vectors. Evaluation demonstrated that the semantic descriptors outperformed the other features in both the precision and recall. However, the highest accuracy was obtained when a combination of all the features was used for retrieval.

Unsupervised classification was used to index heterogeneous information (in the form of wavelets [116] and semantic text data) on decision trees in [117]. A committee of decision trees was used to ensure that individual attributes (either text or image features) were not weighted too highly. A boosting algorithm was applied to reduce the tendency of decision trees to be biased towards larger classes. The proposed algorithm achieved an average precision at five retrieved items of about 79% on a retinopathy data set, and of about 87% on a mammography data set. Without boosting, the results were lower: about 74% and 84% for the retinopathy and mammography data sets, respectively. The study also demonstrated that the approach was robust to missing data with a precision of about 60% for the retinopathy data when less than 40% of the attributes were available in the query images.

Similarly in [118], wavelets were fused with contextual semantic data for case retrieval. A Bayesian network was used to estimate the probability of unknown variables, i.e., missing features. Information from all features was then used to estimate a correspondence between a query case and a reference case in the data set, again using the conditional probabilities of a Bayesian network. An uncertainty component modelled the confidence of this correspondence. The highest precision was achieved when using all features, though the Bayesian method alone outperformed Bayesian plus confidence information on a mammography data set. On the retinopathy data set, the highest precision was achieved by Bayesian plus confidence component.

3.3.4 Retrieval from Diverse Data Sets

The diverse nature of medical imaging means that CBIR capabilities must have the capacity to differentiate between modalities when searching for images. This problem has been taken up by the medical image retrieval challenge at Image-CLEF. Participants submit retrieval algorithms that are evaluated on a large diverse medical image repository [162]. Overviews of submissions to the Image-CLEF medical imaging task can be found in [152–154]. In 2006, Liu et al. [155] proposed two methods for solving this retrieval challenge. The first method used global features such as the average grey-levels in blocks, the mean and variance of wavelet coefficients in blocks, spatial geometric properties (area, contour, centroid etc.) of binary ROIs, colour histograms and band correlograms. The second method divided the image into patches and used clusters of high dimensional patterns within these patches as features. Using multi-class SVMs they were able to achieve a mean average precision of about 68% when using visual features.

Tian et al. [163] used a feature set consisting of local binary patterns and the MPEG-7 edge histogram to compare effect of dimensionality reduction using principle component analysis (PCA); the classification was performed using multiclass SVMs. The accuracy of the dimensionally reduced feature set (80.5% at 68 features) was not very different from the accuracy using all features (83.5% at 602 features). Indeed, the highest accuracy was achieved by the feature set falling between these two extremes (83.8% at 330 features).

Recognising that the categorising of diverse images by modalities is essential to support effective retrieval, Rahman et al. [156] proposed a method for the automatic categorisation and pre-filtering of the search space. The authors reduced the semantic gap by associating low-level global image features with high-level semantic categories using supervised and unsupervised learning via multi-class SVMs and fuzzy c-means (FCM) clustering. The retrieval efficiency was increased by using PCA to reduce the feature dimension while the learned categorisation and filtering reduced the search space. Experiments on the ImageCLEF medical data set showed that pre-filtering resulted in higher precision and recall than executing queries on the entire data set.

In a similar approach, the associations between features in MPEG-7 format and anatomical concepts in the University of Washington Digital Anatomist reference ontology were used to annotate new, unlabelled images [158]. The most similar images, based upon feature distance, were retrieved from the data set on the basis of feature similarity. The semantic annotation for the unlabelled image was derived from the annotations of the similar images. Experiments on the Visible Human data set [164] demonstrated that their retrieval and annotation framework achieved an accuracy of about 93.5%.

3.3.5 Multiple Images and Modalities

A recent study [157] proposed the use of multiple query images to augment the retrieval process. These images were of the same modality: microscopic images of cells. Texture and colour features were used in a two-tier retrieval approach. In the first tier, SVMs were used to classify the major disease type (similar to the approach used by [119]). The second tier was further subdivided into two levels: the first level found the most similar images, while the second tier ranked individual slides using a nearest-neighbour approach for slide-level similarity. The slide-level similarity was weighted according to the distribution of the disease subtypes appearing on the slide and the frequency of that subtype across the entire data set. The method achieved a classification accuracy of 93% and 86% on two separate disease types.

Zhou et al. [159] presented a case-based retrieval algorithm for images with fractures. The algorithm combined multi-image queries consisting of data from different imaging modalities to search a repository of diverse images. The cases in the repository included x-ray, CT, MR, angiography, and scintigraphy images. The cases were represented by a bag of visual keywords and a local scale-invariant feature transform (SIFT) [165] descriptor. Retrieval was achieved by calculating the similarity of every image in the query case with every image in the data set to find the set of most similar images (for a particular image in the query case). The list of all similar images was then reduced to a list of unique cases in the data set. Three feature selection strategies were evaluated and it was demonstrated that feature selection based on case offered the best performance and stability.

The studies described earlier in this section operated on multiple images or multiple modalities but were not designed to retrieve multi-modality images that were acquired on a combined scanner, such as PET-CT or PET-MR. The coalignment of the different modalities in these images offers opportunities for search based on complementary features in different modalities and spatial relationships between regions in either modality.

While clinical utilisation of co-aligned PET-CT has grown rapidly [39,44], few studies have investigated PET-CT CBIR. The only research on this subject was either conducted by our collaborators [124–130] or is described in this thesis [131– 140]⁵. Kim et al. [124] presented a PET-CT retrieval framework that enabled a user to search for images with tumours (extracted from PET) that were contained within a particular lung (extracted from CT) using overlapping pixels. The study introduced the capability to search for tumours by their location or size. Song et al. [125] presented a PET-CT retrieval method using Gabor texture features from CT lung fields and the SUV normalised PET image. Experiments showed that the method had higher precision than approaches that used traditional histograms and Haralick texture features. A scheme for matching tumours and abnormal lymph nodes by pairwise mapping across images was presented in [128]. A weight learning approach using regression for feature selection was presented in [130]. While the algorithms were restricted to thoracic images they showed promise for adaptation to whole body images.

3.4 Summary of Gaps

A number of approaches in the literature have been validated for different image modalities and clinical applications (breast cancer, spinal conditions, etc.). The

 $^{^5\}mathrm{These}$ works will not be discussed in this section as their contents are integrated throughout this thesis.

multiplicity of 2D CBIR research has led to many 2D approaches being applied to images with higher dimensions, e.g., the representation of volumetric images through the use of key slices.

The ImageCLEF medical retrieval task has encouraged research into retrieval from diverse data sets. The CBIR technologies developed as part of the task are well positioned to tackle the challenges in clinical environments where a variety of image modalities are acquired. In particular, the ImageCLEF task has led to the development of methodologies for classifying image modalities based on features. In past years, most of the images in the ImageCLEF medical data set were inherently 2D or 2D constructions of multidimensional data. The data set is expanding to include a greater variety of images (including combined modalities in one image), to inspire further research into the retrieval of such data⁶.

Similarly, the VISCERAL project [166] is a new initiative that aims to provide 10 TB of medical image data for research and validation. In particular the project intends to hold challenges that exploit the knowledge stored in repositories for the development of diagnostic tools. The VISCERAL data set will contain two annotation standards: a *gold corpus* annotated by domain experts and a *silver corpus* annotated by deriving a consensus among research systems developed by challenge participants.

The use of non-image features to complement image features has been widely investigated because all patients have some associated textual data, such as clinical reports and measurements. Metadata such as DICOM tags also provide useful information about the image acquisition protocol, e.g., patient weight, radiotracer dose, etc., and these can also be used to enhance the retrieval process. It has been demonstrated that combining visual features together with text data improves the accuracy of the search but further research is necessary to make the

⁶The variety of data included in the ImageCLEF medical challenges for 2013 can be viewed at http://www.imageclef.org/2013/medical.

contribution of this combination statistically significant [150].

Despite the progress in medical CBIR, there are still several gaps in the stateof-the-art. The Workshop on Medical Content-Based Retrieval for Clinical Decision Support at MICCAI [9], one of the premier forums for medical CBIR research, and several other medical CBIR reviews [27], have identified several areas for investigation. In the following subsections, we detail specific areas for future research that should be pursued to improve CBIR capabilities for modern multi-dimensional and multi-modality medical image retrieval from repositories containing a diverse collection of data.

3.4.1 Multi-dimensional Image Processing and Feature Extraction

Multi-dimensional images are now acquired as a routine part of clinical workflows. However, despite the prevalence of volumetric images (CT, PET, MR, etc.) and time-varying images (4D CT, dynamic PET and MR), some medical CBIR algorithms adopt key slices to represent the entire set of multi-dimensional image data. While this has proven effective in some scenarios, it is highly dependent on the selection of appropriate key slices; manual selection is subjective. In applications where key slices are still viable, subjective selection can be avoided by using a selection algorithm trained by unsupervised learning, as in [167]. In other cases, the use of key slices may not be possible as it may sacrifice spatial information, such as clinically relevant information (a fracture, multiple tumours, etc.) that is spread across multiple sites and slices. Multiple key slices, as in [129, 167], become less viable in cases where the disease potentially spreads across the body, e.g., lymphoma. As such, it would be advantageous if future medical CBIR studies do not rely on key slices but are optimised to operate directly on the rich multi-dimensional image data acquired in modern hospitals. The direct use of multi-dimensional images will require the integration of image processing techniques (compression, segmentation, registration, etc.) that are optimised for such images. The trend towards using local features in generic CBIR [63] indicates that the development of accurate segmentation algorithms will become critical for the development of ROI-based CBIR solutions. The efficiency of some existing algorithms will also need to be optimised for realtime operation. As an example, a recent adaptive local multi-atlas segmentation algorithm [168] extracted the heart from chest CT scans with a mean accuracy of approximately 87% within 30 minutes; such processing times are not feasible for rapid data access without further optimisation.

The curse of dimensionality has always been an issue for medical CBIR algorithms and remains relevant as algorithms are developed for modern medical images. Feature extraction and selection algorithms will need to form a core component of retrieval technologies to ensure that indexing and retrieval can be performed in an efficient manner. Methods that extract multi-dimensional local features from every pixel are no longer feasible for volume and types of images routinely acquired in modern hospitals.

The increasing clinical utilisation of multi-modality images offers the opportunity to derive complementary information from different modalities, the fusion of which will provide extra multi-dimensional features that may not be available from a single image type. The registration of viewpoints and ROIs between imaging modalities will be a key aspect of such CBIR systems, particularly for the extraction of relational features, tumour segmentation given anatomical priors, and fused visualisation. Multi-modality PET-CT and PET-MR scanners inherently provide co-alignment information, allowing the relations between function (tumours) and anatomy (organs) to be emphasised. Future studies should make full use of these features by defining similarity in terms of features from both modalities. In addition, useful indexing features can potentially be extracted from the relationships between ROIs in different modalities. Feature selection algorithms will need to examine the balance between features from individual modalities, as well as relationship features between modalities.

3.4.2 Interpretation and Visualisation

Image retrieval tasks are often carried out for a particular purpose. In medicine, these purposes can include evidence-based diagnosis, physician training, or research. As such, an effective method of showing the images to the user is a critical aspect of CBIR systems. However, there has been limited investigation into methods for the effective interpretation and visualisation of retrieved images, with most studies focusing on improving retrieval accuracy and speed. This not only limits the ability of researchers to evaluate the clinical relevance of their work but also makes it more difficult for clinical acceptance of CBIR as a tool that can be used for daily patient diagnosis and management. This issue exists in general CBIR research as well, where the gap in experiential multimedia search hinders the ability of users to explore and understand the retrieved images [61].

Existing research works that address these problems are often 2D or key slice CBIR systems, such as [169] for non-medical images. The introduction of multidimensional and multi-modality data introduces new visualisation challenges. CBIR systems need to have the capacity to display multiple volumes or timeseries (one for each retrieved image), as well as fusion information in the case of multi-modality images. The systems need to optimise hardware use especially when volume rendering is being used. A number of human factors also need to be considered to enable interpretation of visualised data by users [96]. The visualisation should exploit the retrieval process to demonstrate *why* the retrieved images are relevant.

The development of effective user interfaces is an area of increasing interest, especially if the CBIR systems are to be trialled in clinical environments. User interface guidelines for search applications should be followed to ensure that users are able to easily integrate the CBIR system into their clinical workflow [95]. Context-aware multi-modal search interfaces, such as [170], should be pursued to give users the flexibility to overcome the sensory and semantic gaps.

3.4.3 Inclusion of Clinical Context

As mentioned by Müller et al. [9], the majority of CBIR research is evaluated purely in non-clinical environments. Closer communication is needed with clinical staff to ensure that medical CBIR research has outcomes that are relevant to healthcare. Clinical staff should be involved in the design of CBIR systems; medical specialists should be consulted especially if a domain-specific paradigm [63] is being adapted. However, collaborative projects are hampered by the need for physicians to actively treat patients, leaving limited time for frequent feedback.

It has been indicated that there is a lack of effective representations of medical content using low-level image features. Medical CBIR research could be vastly improved if the clinical context is taken into account, either from direct feedback from physicians, or by using clinical literature as a tool to design CBIR algorithms in line with clinical knowledge. Disease staging and classification schemes, such as those for cancer [171, 172], provide contextual information that can be used to optimise medical CBIR systems based on the guidelines used by physicians. Furthermore, the integration of medical terminology in ontologies such as RadLex [151] and the Unified Medical Language System [173] by learning correspondences between image features and text labels should also be investigated for the case of multi-dimensional images.

The benefits and drawbacks of current algorithms in clinical environments should also be examined. The lessons learnt from such studies will benefit the design and optimisation of algorithms in future CBIR investigations. This will lead to CBIR frameworks that are practical, usable, and valuable for the intended task.

3.4.4 Data Sets for Evaluation

Most medical CBIR research is evaluated on closed data sets that are available only to the authors. This makes it difficult to compare different CBIR algorithms. The creation of standardised data sets, such as the ImageCLEF medical data set, should assist in alleviating this problem. Other data sets that can be used for this purpose include the Lung Imaging Database Consortium (LIDC) [174], the OASIS data set of MR brain images [175], and the new VISCERAL data set [166].
Chapter 4

A Graph Primer

This chapter provides an overview of graphs, graph-based representations of data, and graph algorithms that are the basis for later sections of this thesis.

4.1 Definitions

A graph G = (V, E) is a pair of sets V and E, where $V = \{v_1, v_2, \ldots, v_n\}$ and $E = \{e_1, e_2, \ldots, e_m\}$. The elements of V are called the vertices or nodes of G. The elements of E are known as the edges or arcs of G. Every edge $e_k = (v_i, v_j) = v_i v_j$ is a pair of vertices $v_i, v_j \in V$, where $1 \leq i, j \leq n, i \neq j$, and $1 \leq k \leq m$. The vertices v_i and v_j are called the endvertices of e_k ; v_i and v_j are said to be adjacent to each other and incident to e_k . The order of a graph, denoted by |G|, refers to the number of vertices in G, i.e., |G| = |V| = n. The size of a graph, denoted by ||G||, refers to the number of edges in G, i.e., ||G|| = |E| = m.

In this thesis, graphs are visualised by depicting vertices as circles and edges as lines that connect two vertices. The graph in Figure 4.1 has an order of 5 and a size of 6. Its vertex set consists of the vertices A, B, C, D, and E. The vertices C and D are adjacent while the vertices D and E are not. The vertex C is an endvertex of the edges BC, CD, and CE, and is incident to all of them.



Figure 4.1: A graph with five vertices.



Figure 4.2: Two subgraphs of Figure 4.1.

A graph G' = (V', E') is a *subgraph* of G if $V' \subseteq V$ and $E' \subseteq E$. That is, G' is a subgraph of G if its vertex set (V') and edge set (E') are subsets of the vertex set and edge set of G. Figures 4.2(a) and 4.2(b) are two subgraphs of Figure 4.1. The vertex E, and the edges CD, and CE have been removed from Figure 4.1 to form Figure 4.2(a). Similarly, the vertex E and the edges AD, BD, and CE have been removed to form Figure 4.2(b).

A complete graph is a graph where all vertices are pairwise adjacent, i.e., there exists an edge between all vertices in the graph. A *clique* is a subgraph that is also a complete graph. An example of a clique of Figure 4.1 is the subgraph formed by the vertices A, B, and D, and the edges AB, AD, and BD.

Two graphs G = (V, E) and G' = (V', E') are said to be *isomorphic* if the elements of V can be mapped to the elements of V', and vice versa, while preserving the structure of the graphs. More specifically, if G and G' are isomorphic then there exists some bijective mapping $\varphi : V \to V'$ where $\forall e = (x, y) \in E \Leftrightarrow e' = (\varphi(x), \varphi(y)) \in E'$. That is, for every pair of adjacent vertices in G there



Figure 4.3: Isomorphism and subgraph isomorphism. (a) is isomorphic to Figure 4.1; the blue subgraph in (b) is isomorphic to Figure 4.1.

is an equivalent pair of adjacent vertices in G'. The graph G' is subgraph isomorphic to G if there is some subgraph of G' that is isomorphic to G, i.e., a subset of adjacent vertices in G' has an equivalent pair of adjacent vertices in G. Figure 4.3 depicts these concepts. Figure 4.3(a) is isomorphic to Figure 4.1 under the mapping $\varphi(A) = 1, \varphi(B) = 2, \varphi(C) = 3, \varphi(D) = 4$, and $\varphi(E) = 5$. Figure 4.3(b) is subgraph isomorphic to Figure 4.1 because the subgraph in blue is isomorphic to Figure 4.1 under the same mapping as given before.

4.2 Representing Structures with Graphs

Pattern recognition approaches generally fall into one of three categories: syntactical, statistical, and structural [176]. Syntactical approaches encode data as elements of a grammar, and discriminate between different classes of objects using formal language theory [177]. As discussed in Section 1.1, the volume of data and the need for analysis by domain experts make syntactical approaches largely unfeasible for medical image pattern recognition.

In statistical approaches, an object or pattern is represented by a feature vector $\mathbf{f} = (f_1, f_2, \dots, f_n) \in \mathbb{R}^n$. Each $f_i, 1 \leq i \leq n$, is a real-valued measurement of a particular feature of the object or pattern. A large number of mathematical operations can be efficiently computed in a vector space and this has resulted in wide variety of algorithms for statistical pattern recognition [178]. However, vectors are inherently unable to encode relationships between different parts of an object or pattern. Furthermore, the feature set for any given application must remain fixed, and all vectors must be of equal length, regardless of the complexity of the patterns or objects described by individual vectors.

Structural approaches represent objects or patterns as graphs (or special cases of graphs like trees or strings). Components of the objects become graph vertices while spatial, temporal, and conceptual relationships between these components are represented by the edges of the graph. Unlike vectors, graphs are not constrained to a fixed length and can adapt to the complexity of the pattern being represented; the order of the graph can change to match the number of components in the pattern while the size of the graph can be altered according to the relationships in the pattern. However, tools for graph-based pattern recognition are not as rich as for statistical pattern recognition largely due to the computational complexity of graph-based algorithms [179].

Several methods have been invented to bridge the gap between the statistical and structural approaches. The most general of these is the attributed relational graph (ARG), an encoding of pattern data that combines vectors and graphs [180], which is considered the standard definition of graphs in pattern recognition today [176]. An ARG is a graph $G = (V, E, \alpha, \beta)$. The function $\alpha : V \to L_V$ is a vertex labelling function, on a set of possibly infinite vertex labels L_V . Similarly, $\beta : E \to L_E$ is an edge labelling function, on a set of possibly infinite edge labels, L_E . In essence, every ARG vertex and edge is labelled with a feature vector that describes the properties of the object or the relationship, respectively for vertices and edges. These feature vectors are also known as the attributes of the ARG. ARG representations are capable of indexing almost any feature as vertex and edge attributes by expanding the label sets L_V and L_E [121].



Figure 4.4: Constructing a graph representation of a pattern.

Figure 4.4 depicts a pattern (Figure 4.4(a)) and one possible ARG representation (Figure 4.4(b)) of that pattern. Each vertex of the graph represents the object of the same colour. Edges have been created between objects that are overlapping or touching. Each vertex has a single attribute Λ for the area of the object (measured in square centimetres), while every edge has a single attribute δ for the spatial distance between the location each object's centre (measured in centimetres). That is, $\alpha = \{\Lambda\}$ and $\beta = \{\delta\}$, where $\Lambda = area (ROI)$ and $\delta = distance$ (ROI1, ROI2); in this example $L_V = \{\mathbb{R}_{>0}\}$ and $L_E = \{\mathbb{R}_{\geq 0}\}$. While we have only depicted one attribute for both vertices and edges, they do not need to have the same number of attributes. The unit of measurement for the attributes can vary depending on the application domain; while we have used centimetres in our example, it is equaly valid to use pixels, voxels, or millimetres as the situation demands.

There are various forms of ARGs. A regional adjacency graph (RAG) represents ROIs and the relationships between ROIs that share a common border [181], e.g., Figure 4.4(b). The curvature tree represents the curvature information of ROIs and the relationships between ROIs that are bounded by or contained within others [82]. The hierarchical attributed regional adjacency graph (HARAG) is an RAG with extra relationships between larger ROIs and their smaller constituent ROIs [81].

4.3 Measuring Graph Similarity

The measurement of graph similarity is a question of the degree to which the structures of the graph are the same. Graph isomorphism and subgraph isomorphism are specialised forms of this measurement and have a binary outcome, i.e., two graphs are either isomorphic (or subgraph isomorphic) or they are not. The following subsections discuss two methods for measuring graph similarity: the traditional approaches based on graph edit distance, and modern approaches based on the concept of kernel machines.

4.3.1 Graph Edit Distance

The graph edit distance $D_{ge}(G, H)$ is a method for measuring the similarity of two graphs, G and H, by measuring the amount of distortion to morph one graph into the other [182]. This process is also known as *error-tolerant graph matching*. The graph edit distance calculation measures the cost to make the graph structures isomorphic as well as the cost to make vertex and edge attributes the same. The calculation of D_{ge} requires the definition of a sequence of graph transformation operations (such as vertex insertion, deletion, etc.), and an associated cost for each of these operations. The algorithm for calculating D_{ge} attempts to find the series of operations that has the minimum cost. The cost becomes the degree of dissimilarity (with a cost of 0 implying the two graphs are isomorphic and have the same attributes).

Let $o = \langle g, h \rangle$ be an edit operation, where g is a vertex of G and h is a vertex of H or, alternatively, g is an edge of G and h is an edge of H. When g and h are not null (\emptyset) then o is a substitution operation. When g or h is null then o is insertion or deletion. Using the terminology for graph isomorphism, substitution can be defined as $\varphi(g) = h$ while insertion or deletion can be defined as $\varphi(g) = \emptyset$ or $\varphi(h) = \emptyset$. The graph edit distance can then be defined as:

$$D_{ge}(G,H) = \min_{(o_1,o_2,\dots,o_n)\in\Omega} \sum_{i}^{n} d(o_i)$$
(4.1)

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 . For ARGs the cost function $d(\cdot)$ can be derived from the attributes that are indexed on the vertices and edges. As an example, Petrakis et al [183] adapted D_p (Equation 3.1) as the cost for the various different transformation operations.

However, the flexible nature of the graph edit distance calculation is also its downfall, especially in application to unconstrained graphs. The D_{ge} calculation works by finding the cost of every sequence in Ω , and then selecting the sequence that has the minimum cost overall. Larger graphs will have a larger number of sequences in Ω . As such, brute-force algorithms for calculating D_{ge} have very high time and space complexities. Restricting the calculation to special classes of graphs, such as trees can improve the efficiency of the process [184]. Formulating the process as a vertex labelling problem enables the use of relaxation labelling techniques [185].

Most graph edit distance calculations are formulated as tree search problem. One such search tree is depicted in Figure 4.5. In this example, G (Figure 4.5(a)) is being compared with H (Figure 4.5(b)). The nodes of the search tree (Figure 4.5(c)) represent edit operations between the vertices of G and H (and \emptyset). The path from the root (*) to a leaf node is a unique sequence in Ω .

Figure 4.5(d) shows the expansion of one branch of the search tree (marked with a double line in Figure 4.5(c)). In the blue path, vertex A is substituted with vertex Y at the first node. The second node substitutes vertex B to vertex Z. Edge AB is deleted as there is no edge incident to Y and Z in H. At the third



Figure 4.5: Graph edit distance search tree for the graphs given by (a) and (b). The search tree (c) represents all possible combinations of vertex edit operations between the two graphs. Every branch from the root to leaf (such as the blue one given in (d)) is a unique sequence of edit operations.

node, vertex C is deleted, resulting in the deletion of edge AC. In the leaf node, vertex D is substituted with vertex X. This causes edge BD to be substituted with edge XZ and the insertion of an edge between A and D to correspond to edge XY. The blue path would have a total cost that aggregates the substitution costs of A and Y, B and Z, D and X, and B and XZ; the deletion costs of AB and AC; and the insertion cost of XY (as AD).

One of the most notable algorithms for computing the graph edit distance is the A^{*} algorithm [186], a best-first algorithm that constructs the search tree dynamically. The root of the tree is the starting point, and every branch represents a decision point for vertex operations. As such, every path from the root to a leaf represents a different possible sequence of transformation operations. A heuristic function at every search tree node estimates the cost from that node to a leaf, allowing the algorithm to determine whether a particular search space is worth exploring. However, even with the heuristic function, the A^{*} algorithm has at worst an exponential complexity.

Two adaptations of the A^{*} algorithm were proposed in [187]. The path length A^{*} approach introduces an additional weighting to the A^{*} algorithms optimal path and heuristic cost functions to avoid expanding the search tree when a node with a significantly large cost is encountered during vertex mapping. Faster search is enabled because similar graph elements have very cheap transformation costs. The beam search A^{*} algorithm approach [187] 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 operation by limiting the number of sequences that have to be evaluated. While the beam search algorithm is suboptimal in accuracy, it was demonstrated that the beam search approach was almost as accurate as a brute force approach when using large beams [187]. When applied to a classification task, the suboptimality of the algorithm resulted in an increase in inter-class differences while intra-class differences were not strongly affected, demonstrating that the algorithm was appropriate for ranking objects images based on similarity of their classes. Another solution for measuring graph similarity is the Hungarian method for solving the assignment problem [188, 189], which is a polynomial time solution that operates entirely on vertex attributes. However, this method produces suboptimal graph similarities because it ignores edge attributes. Thus it is not viable for graph similarity except when adapted to graphs where the edges encode a specific binary relationship, e.g., inclusion in [82].

Petrakis et al [121] experimentally demonstrated that measuring ARG similarity with a brute-force graph edit distance calculation was more precise than other approaches such as the Hungarian method when searching for images based on spatial similarity. They also demonstrated that the ARG representation allowed for more precise retrievals compared to other graph-like structures such as strings. However, this precision came at a cost; there was a trade-off between accuracy and computational performance.

4.3.2 Graph Kernels

Kernel machines [190, 191] address pattern recognition problems by solving them in a related vector space instead of the original space. Graph kernels enable graph similarity measurement by mapping all graphs into a vector space where the rich mathematical tools for the domain of vectors can be applied [192]. The difficultly lies in finding a vector space mapping that preserves the structural similarity of graphs. A key result of graph kernel machines is that an implicit embedding of the entire graph space to a vector space can be formulated from the definition of another graph similarity measure [192]. It was shown in [193] that deriving a general graph kernel on the basis of subgraph isomorphism is a computationally intractable problem.

A number of graph kernels related to graph edit distance were described in [194]. Diffusion kernels are trained from a subset of the entire graph space. The training process turns a matrix of pairwise graph edit distances (between the graphs in the subset) into a diffusion kernel by considering not only how similar two patterns are, but also how many similar patterns they have in common.

Convolution kernels [194] are calculated on the basis of decomposed graphs, an ordered sequence of vertices, and the edges connected to those vertices. These decompositions can be interpreted as edit distance paths (paths from the root to the leaf node of the search tree). The kernel over all possible, compatible decompositions determines the similarity of corresponding elements of the edit distance path. The convolution kernel is inherently inefficient because the number of decompositions grows exponentially with the order of the graphs.

Random walk kernels [194] define the similarity of graphs by comparing random walks. Two graphs are considered similar if they share a large number of matching random walks. The random walk kernel essentially evaluates the local similarity of parts of two graphs. The random walk kernel was enhanced by the graph edit distance to render classification performed with the kernel more robust to noise. The modification removed from the kernel computation those nodes that violated the optimal vertex correspondence identified by the minimum cost transformation path. However, the computational complexity of this kernel is comparable to the graph edit distance.

A set of experiments compared these kernels and the nearest neighbour approach in conjunction with an SVM on various data sets [194]. All three kernels showed statistically significant improvement of the nearest neighbour approach on a data set containing line drawings of capital letters. The diffusion and convolution data sets were significantly better than the nearest neighbour when classifying a data set of images. The convolution and edit distance modified random walks kernel outperformed the nearest neighbour approach on a finger print data set.

Graph kernels enable the use of vector space algorithms and tools (such as SVMs) on graphs. However, because the derivation of the aforementioned kernels

is still generally dependent upon the graph edit distance or is similarly computationally expensive, it is important to first establish an edit distance algorithm before considering solutions that use kernel machines.

Chapter 5

CBIR Design

This chapter represents the start of the methodology section of the thesis. We begin by describing the design of our proposed CBIR system for multi-modality medical images, and specifying the manner in which our design contributes to solving the "gaps" in the state-of-the-art.

5.1 Overview

As described in Chapter 3, most CBIR technologies require the following components: a method for extracting features from images; a representation for indexing these features; a technique by which the indexed features, and consequently their images, can be compared; and an interface that communicates the findings back to the user who initiated the search. Existing approaches for systems and components have already been discussed in Sections 3.2 and 3.3; as such this section will only describe deviations from the norm that are motivated by our multi-modality image domain.

Our retrieval framework is designed around the following notion of image similarity: "an retrieved image is relevant to a query image if the location of tumours relative to anatomy within both images is similar". That is to say, we defined a *true positive* retrieved image as one where the localisation of tumours was shared with the query image. This criterion for image similarity was derived from clinical guidelines for cancer staging and disease staging [171,172,195], which classifies the disease according to tumour characteristics and the relationships of tumours with surrounding anatomical structures.

In particular, our proposed design requires the following modifications to the standard CBIR framework:

- Separate segmentation and feature extraction algorithms for the different modalities of multi-modality images. This enables optimised ROI detection, and the extraction of feature sets that are better suited to a particular image modality.
- 2. The inclusion of cross-modality registration to enable extraction of relationships between ROIs from different modalities.
- 3. A single representation that indexes image features from different modalities, as well as relationships between modalities.
- 4. A feature normalisation scheme that can scale different types of features from different modalities into the same range. This must be done to ensure that no feature inherently biases similarity measurement.
- 5. A similarity measurement algorithm that is capable of deriving a single similarity value on the basis of information from different modalities. The similarity measurement must take into account all image features, as opposed to comparing similarity by modality and then fusing the results together.
- 6. A method for visualising and interpreting the retrieved results. This means the implementation of a user interface that displays complete volumes (rather than key slices) of the individual images, as well as the fused images, thereby

enabling users to visually verify image similarity on a per modality and fused basis.

In particular, our representation and similarity measurement algorithm will integrate clinical knowledge to ensure that our CBIR technique is relevant to the clinical context. More specifically, our representation uses the geometric and topological spatial features to constrain tumours to their proximal (spatiallynearest) organs.

5.1.1 Assumptions

Our multi-modality CBIR design is predicated on the following assumptions:

- The images only consist of two modalities, each of which present both common and complementary information. In particular we assume that anatomical data is provided by one modality (CT), and functional or pathology data from another (PET).
- 2. Cross-modality registration is performed by the scanner. Hardware coalignment is a common feature of modern multi-modality scanners. For the purpose of simplicity, we disregard registration errors in this thesis.
- 3. ROIs may be extracted using any segmentation algorithm. We assume that the framework need not use segmentation algorithms that delineate specific structures. As such, the ROIs are only labelled by their modality.
- 4. Each ROI is a collection of pixels (2D) or voxels (3D) belonging to either an anatomical structure or a tumour.

5.2 Framework

Figure 5.1 shows the major components of our CBIR framework. The components marked in grey indicate the main contributions of this thesis; later chapters describe them in greater detail. The data flow for the offline indexing process is depicted by the pink arrows. The green arrows indicate the flow of data during the real-time querying process. These processes are described in detail in the following subsections.

5.2.1 Data Set Indexing

The data set indexing process converts the images stored in a repository to graphs that represent the content (features) of those images. The entire process can be done offline prior to any query being run.

- The DICOM images in the PACS are converted to fast-loading TIFF images [196], which are stored on disk. The fast-loading TIFF stacks allow the image processing algorithms and user interface to efficiently load images during processing or visualisation.
- Segmentation is performed to select ROIs from the images. If necessary, more than one segmentation algorithm may be used. The ROIs from both modalities are stored on disk.
- ARGs are constructed from each of the ROIs. These graphs are stored on disk. See Section 6.2 for more details.
- 4. The index is normalised before being used for a query to reduce feature bias during similarity measurement. See Section 6.4 for more details.



Figure 5.1: Proposed CBIR framework. The pink arrows show the process during offline indexing; the green arrows show the process flow when a real-time query is being performed. The grey boxes indicate elements of the framework that are described in detail, in other sections of this thesis.

5.2.2 Querying

The query process converts a query image to a graph, and then compares the query graph with the graphs created during the indexing process. This process is carried out in real-time.

- 1. The user inputs a query image. A graph is created for this image following the same process used during data set indexing, except that the query graph is not stored in the index. See Section 6.2 for more details.
- 2. The query graph is normalised. See Section 6.4 for more details.
- 3. The query graph is compared to the graphs in the index. The output of the similarity calculation is used to construct a ranked list of the most similar images. See Section 6.6 for more details.
- 4. The user interface presents the query image and the retrieved images to the user. The clinical reports and segmented ROIs provide supplementary information that assists the user in interpreting the retrieved images. See Chapter 8 for more details.

5.3 Design Justification

Our definition of image similarity arose from clinical literature describing cancer staging and classification [171,172,195]. These schemes generally characterise different disease stages according to the proximity, size, and relationships between tumours to surrounding anatomical structures. For example, under the TNM classification scheme for lung cancer [171], a T1 primary tumour is surrounded by the lung or visceral pleura and is always ≤ 3 cm in size while a T2 primary tumour is always > 3cm in size (but ≤ 7 cm) and can invade the visceral pleura or involve the main bronchus; meanwhile, T3 primary tumours may invade other structures such as the chest wall, diaphragm, or mediastinal pleura. Similar location based criteria are applied for the categorisation of lymph node involvement (the N-descriptor) and metastatic disease (the M-descriptor). Another example is provided by the Ann Arbor staging system for lymphoma [195]: Stage III includes involvement of lymph nodes on both sides of the diagram and may be accompanied by involvement of the spleen, while Stage IV indicates diffuse or disseminated diseases, such as involvement of the liver¹. Our definition of image similarity, based as it is on the properties of tumours and their location relative to anatomy, therefore mimics the geometric and topological characteristics used for clinical staging and classification.

We chose a graph-based representation for multi-modality images due to a number of factors. Firstly, graphs are more capable at representing structural image information than vector-based approaches because they are not-constrained to a predetermined fixed size or order [179]. This was particularly important for cases with multiple tumours and organs, each with different characteristics; a graph-based approach enables every element to be represented independently instead of accumulating features. Graph edges allow the spatial arrangement of ROIs to be quantitatively characterised unlike directional matrices [85], which are limited to dividing relative location into a set number of bins. In addition, unlike complex strings [86] graphs can inherently accommodate 3D data. Graphs have also been demonstrated to be more accurate than other approaches when retrieving images based on the spatial arrangements of objects [121] meaning a graph-based representation was ideal for our similarity definition.

The major drawback of graph representations comes from the high computational complexity of many graph algorithms, including the most accurate brute

¹Note that for the purposes of brevity, the examples given here for both TNM and Ann Arbor staging have been simplified. In reality, these descriptors often have numerous substages for more specifically identifying the characteristics of the disease.

force methods for calculating the graph edit distance [121, 179] (see also Section 4.3). The number of ROIs in our data sets made brute force approaches impractical. While faster methods are generally less accurate, Neuhaus et al. [187] demonstrated that the beam search approach was almost as accurate as a brute force method when using large beams. In addition, they demonstrated that the beam search algorithm was appropriate for ranking objects images based on similarity of their classes because it increased inter-class differences while intra-class differences were not strongly affected. We therefore adapted the beam search algorithm to enable the comparison of graphs representing multi-modality images.

Chapter 6

Multi-Modality Image Retrieval

In this chapter, we present our method for the graph-based retrieval of multimodality medical images. We first define graphs for representing multi-modality images. We also describe the multi-modality regional and relational image features to be indexed as attributes of these graphs. We then detail a scheme for normalising the indexed features in a modality-specific manner. Finally, we present adaptations to the graph edit distance algorithm that enables the comparison of multi-modality graphs.

6.1 Overview

Figure 6.1 illustrates our graph construction process. During the first stage, ROIs are delineated in each image using segmentation. Different segmentation algorithms are applied to different modalities. In the figure, two tumours (bordered in red) are extracted from the PET image and the lungs (bordered in blue) are extracted from the CT image. Features are then extracted from each ROI. The extracted features include those that occur in only a specific modality. A vertex is then created for every ROI; an ROI's features are assigned as the attributes of its vertex.



Figure 6.1: Multi-modality graph construction process.

Registration information (such as the co-alignment acquired by the scanner hardware) is then used to align different ROIs within the same coordinate space. The co-alignment is used to calculate spatial relational features between two ROIs extracted from different modalities. These relational features are indexed as attributes of the edges between the vertices representing the two ROIs. An edge is created for every pair of vertices, resulting in the construction of a complete graph (Section 6.2.1). Finally, an edge filter constrains the edges using criterion based on the geometric and topologic attributes of the disease classes (Section 6.2.2).

6.2 Multi-modality Graph Definitions

According to Assumption 1 (see Section 5.1.1), the individual modalities provide either anatomical or functional (pathological) information. We therefore divided our vertex set into two, giving us the following definitions.

Let V_A be the set of vertices that represent anatomical ROI. Similarly, let V_P be the set of vertices representing functional ROIs. Let f_A and f_P be the sets of features extracted from the anatomical and functional ROIs, respectively. In addition, let $\alpha_A : V_A \to L_A^j$ and $\alpha_P : V_P \to L_P^k$ be the vertex labelling functions for V_A and V_P , respectively, where L_A and L_P are their respective label sets, $j = |f_A|$, and $k = |f_P|$. Let $\beta_S : E \to L_S^m$ be the labelling function for all edges, with feature set f_S , label set L_S , and where $m = |f_S|$. This definition treats each vertex or edge of a graph as an individual feature vector.

The feature sets f_A , f_P , and f_S are dependent upon the characteristics of the data set, i.e., different multi-modality images will have different features. The features used in this thesis are defined in Sections 6.3.

Given these notations, we defined the graph representations for a given multimodality image I.

6.2.1 Complete Graph

We indexed all features and all relationships extracted from the images using a complete graph. The pairwise adjacency of all complete graph vertices ensured that relationships between all ROIs were preserved and considered equally important. This enabled us to model anatomical variations between patients, e.g., minor differences in the separation of organs, as well as the relationships between



Figure 6.2: Creating a CAPP graph by pruning edges. The pruning function Φ is applied to the complete graph representation of the PET-CT image in Figure 6.1. The green edge is preserved ($\Phi = 1$) because it is incident to two vertices in V_A . The purple edges are preserved because they represent relationships between tumours (V_P) to the nearest anatomical structures (V_A). The dashed edges do not meet either of these criteria and are pruned ($\Phi = 0$).

tumours to all anatomical ROIs.

We defined our multi-modality complete graph as $G_K = (V_K, E_K, \alpha, \beta, I)$ where $V_K = V_A \cup V_P$ is the vertex set, $E_K = \{v_i v_j\} \forall v_i, v_j \in V_K$ where $i \neq j$ is the set of all edges, $\alpha = (\alpha_A, \alpha_P)$ is the combined vertex alphabet, and $\beta = \beta_S$ is the edge alphabet.

6.2.2 Complete-Anatomy Proximal-Pathology Graph

We hypothesised that the complete graph representation could be improved by constraining tumours to the most closely related anatomy by pruning edges that connect tumours to anatomical structures that are not likely to be directly related. This de-emphasised the less relevant information between unrelated structures. For example, if e_1 was an edge between a *lung* tumour vertex and a brain vertex, and e_2 was an edge between the tumour vertex and a lung, then e_1 would be pruned while e_2 would be preserved. We applied this pruning process to the complete graph to obtain the Complete-Anatomy Proximal-Pathology (CAPP) graph. The CAPP graph used the geometric and topological spatial features indexed to constrain tumours to their proximal (spatially-nearest) organs.

We therefore defined our CAPP graph as $G_{CAPP} = (V_K, E_{CAPP}, \alpha, \beta, I)$ where

 $E_{CAPP} \subseteq E_K$. We generate E_{CAPP} by pruning the complete edge set E_K using an indicator function Φ . Let $v_i, v_j \in V_K$ with $v_i \neq v_j$ be the endvertices of an edge $v_i v_j \in E_K$. The function Φ uses the modalities and proximity of v_i and v_j to signal inclusion within E_{CAPP} . We defined Φ as follows:

$$\Phi(v_i v_j) = \begin{cases} 1 & \text{if } v_i \in V_A \text{ and } v_j \in V_A \\ & v_i \in V_A \text{ and } v_j \in V_P \text{ and} \\ 1 & \text{if} & v_i = \underset{x \in V_A}{\operatorname{arg min} \ proximity(xv_j)} \\ 0 & \text{otherwise} \end{cases}$$
(6.1)

where proximity $(v_a v_p) : E_K \to \mathbb{R}$ is a function that obtained the spatial nearness of the regions represented by vertices $v_a \in V_A$ and $v_p \in V_P$ from the edge $v_a v_p \in E_K$. The value returned by proximity $(v_a v_p)$ can be calculated from the features (f_S) indexed on it; in our experiments this was done on the basis of the *md* feature (see Section 6.3 for an explanation of the different features).

Figure 6.2 depicts the creation of the CAPP graph of the PET-CT image in Figure 6.1. 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 V_A representing the spatially nearest anatomical ROI. The dashed edges have been removed to produce the CAPP graph.

Our definition of Φ limited the relationships between elements of V_P and V_A . As such, it constrained tumours to their spatially nearest anatomical structures, thereby creating a representation where tumour localisation was given a greater emphasis compared to the complete graph. Furthermore, by preserving edges incident to two vertices in V_A our graph graph created a clique (a complete subgraph) containing all the vertices representing anatomical structures. This enabled the CAPP graph to represent all relationships between all anatomical structures. This was necessary to model minor variations between individuals because most humans have the same anatomical structures in the same general arrangement, i.e, the structure is generally the same while individual features (e.g. organ size) may be different.

Edges incident to vertices in V_P were pruned during CAPP graph construction because it was not necessary to explicitly encode the relationships between PET ROIs. The information represented by these edges was already implicitly contained by the relationship features on the other edges. For example, the relative location of two tumours could be reconstructed using the relationships between their nearest-anatomical neighbours and the edges of the complete anatomical subgraph.

6.3 Graph Attributes and Image Features

6.3.1 Types of Features

As noted in Section 6.2, our feature set was divided 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) . As in [116, 118], the types of images contained by a data set determined its feature set, e.g., texture features for CT images, SUV features for PET images, area for 2D images, and volume for 3D images, etc. This enabled the use of features that exploited the characteristics of a particular data set.

In this thesis, we only consider three types of features: measurements, angular values, and point sets. Measurements are features that represent a certain property such as distance or size. Angular features are also measurements except that they may consist of multiple numbers that describe different components of a single feature, e.g., the use of two angles (the pitch and yaw) to describe the relative positions of two ROIs in a 3D space. Finally, point sets are used to describe a collection of coordinates, e.g., the boundary of an ROI.

Our graphs are not inherently limited to these types of features. The definition of the ARG and thus our CAPP graph leaves open the possibility that any possible feature can be indexed as a graph attribute [121]. These features need not be numeric in nature and could potentially include text annotations, or even other sub-images. New feature types may require a unique normalisation and similarity calculation methods to those given in Sections 6.4 and 6.6.1, respectively.

6.3.2 Image Features

In this thesis, we indexed the following features as vertex attributes: size (s), boundary (b), length or the maximum distance between two points on an ROI's boundary (l), roundness (r), and tumour homogeneity (th). The maximum (I_{MAX}) , mean (I_{μ}) , and standard deviation (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), the spatial separation of the centroids of two ROIs; relative orientation (ro), the angle between the centroids of two ROIs; relative size (rs), the ratio between the sizes of two ROIs; and minimum distance (md), the minimum separation between two points in different ROIs.

Figure 6.3 shows the difference between the d and md edge features of a relationship between two lung ROIs. The red line joins the centroids of the ROIs; this distance is indexed as the d feature and represents the separation of the centres of mass of the two ROIs. The blue line joins the closest points in either ROIs. It represents the shortest distance between the ROI and is indexed as the md feature.

Figure 6.4 explains our concept of relative orientation. The red line joins the centroids of two lung ROIs, while the blue lines are the xy axis of the image



Figure 6.3: Calculating the distance and minimum distance edge features between two lung ROI. The red line is the distance between the centroids of the two lungs (d feature) and the blue line indicates the shortest distance between a point in the left lung and a point in the right lung (md feature).

centred over the centroid of one of the ROIs. The *ro* feature is the angle θ (or $2\pi - \theta$) made between the axes and the line joining the centroids. Together the *d* and *ro* features represent the location of the centre of mass of an ROI relative to the centre of mass of another.

Several of these features had corresponding 2D and 3D counterparts. Size was represented by area for ROIs in 2D images and by volume for ROIs in 3D images. Similarly, the boundary size was defined as the perimeter of the 2D and the surface area of the 3D ROIs. For 2D images, ro was a single value representing the angle between the x-axis and the semi-major axis of the ROI. For 3D images, ro consisted of two values, equivalent to the pitch and yaw angles directing an observer from one ROI's centroid towards the centroid of the other.

We defined the 2D feature for roundness (or circularity) as a function of the ratio of the lengths of the region's semi-major and semi-minor axes. We defined sphericity, the corresponding 3D feature, as a function of a region's volume (the



Figure 6.4: Calculating the relative orientation edge feature between two lung ROI. The red line joins the centroids of the two lungs and the blue line indicates xy axis of the image. The angle θ between this axis and the line joining the two centroids is used to represent relative orientation.

s feature) and surface area (the b feature), as in [197]:

$$sphericity = \frac{\pi^{\frac{1}{3}} (6s)^{\frac{2}{3}}}{b}$$
 (6.2)

Tumour homogeneity was a measure of the uniformity of tumour pixel intensities. We adapted the voxel neighbourhood homogeneity coefficient given by [38], 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 $P = \{p_1, p_2, ..., p_n\}$ be the set of voxels within a tumour. The tumour homogeneity is then given by:

$$th\left(P\right) = \frac{\sum_{i=1}^{n} \Lambda\left(p_{i}\right)}{n} \tag{6.3}$$

and

$$\Lambda(p) = \frac{1}{1 + \sqrt{\frac{1}{|nd(p)|} \sum_{k=1}^{|nd(p)|} (p_k - p)^2}}$$
(6.4)

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 grey-level co-occurence matrices (one for each unique direction) were used to calculate 3D Haralick texture features [72, 73]. We calculated five wellestablished features from these matrices: entropy (ent), contrast (cont), correlation (cor), energy (nrg) and homogeneity (hmgt).

We extracted point sets (pts) representing the coordinates of all the pixels in a given ROI from data sets where the image acquisition protocols used the same patient orientation (e.g. head-first and supine) for images with the same resolution. We used this to improve anatomical matching by measuring the overlap between anatomy ROIs in the query and data set images, as in [124].

The features for our experimental data sets are given in Section 7.1.

6.3.3 Feature Set Justification

Our choice of features was primarily motivated by the need to represent the geometric and topological attributes used for cancer staging and classification [171, 172,195], such as geometric tumour properties (volume, length, etc.) and tumour location relative to anatomy (distances and angles from other structures). For this purpose, we adapted the geometric and spatial features described in [121] and complemented them with modality specific information, such as SUV for PET or texture for CT.

We did not perform feature selection to find the optimal set of features. Selecting the features for a graph representation is not a trivial task as it must balance structural information as well as the image features indexed on vertices and edges. This is because, 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 structures in the same graph. For example, the I_{MAX} feature may only be optimal if it occurs on a tumour that has particular rd and rs relationships with a mediastinum that contains multiple tumours; the feature may not be optimal for other tumours. To the best of our knowledge, there is no standard toolset for performing such feature selection. We have therefore left this to future work.

6.4 Graph Feature Normalisation

As explained in Section 3.2.2, there is a danger that features with high numerical absolute values may bias the similarity calculation over features that may have relatively smaller values. This issue can be resolved by normalising features or the differences between features to a standard range thereby ensuring that each feature contributes equally to the distance function. Aksoy et al. [87] demonstrated that normalising features to a fixed range improved the discriminatory capabilities of similarity measures in image retrieval applications.

One technique for normalising feature vectors linearly scaled a feature value to a random variable with zero mean and unit variance, ensuring that 99% of all values were normalised to the range [0, 1]; a ceiling and floor operation was applied guarantee that all normalised values fell within this range [87]. While the vertices and edges of our multi-modality graphs are essentially feature vectors this approach could not be directly applied to all the features indexed on our graphs because of our different feature sets and feature types.

The following subsections provide a summary of our feature normalisation technique for each of the feature types used in our experiments. 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]. Detailed algorithms for performing the feature normalisation are provided in Appendix A.

6.4.1 Measurements

Because measurements were features that represented a certain property it is possible to scale the features using the technique in [87]. However, in our multimodality case it is possible for the range of features to vary widely across different modalities (e.g., tumour volume can be much smaller than organ volume). As such, this normalisation was applied separately for different modalities, e.g., anatomical volume was 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 data set. The normalised value \tilde{x} of x was determined the following function [87]:

$$\tilde{x} = \frac{\left(x - \mu_f\right)/3\sigma_f + 1}{2} \tag{6.5}$$

6.4.2 Angular Values

While angular features (such as relative orientation) are also measurements, the circular nature of the measurements makes them difficult to normalise by linear scaling as in Section 6.4.1. 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 value θ as functions of its *sine* and

cosine [121]. The normalised value $\tilde{\theta}$ was a pair of values:

$$\tilde{\theta} = \left\langle \frac{\sin\left(\theta\right) + 1}{4}, \frac{\cos\left(\theta\right) + 1}{4} \right\rangle \tag{6.6}$$

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

6.4.3 Point Sets

We did not normalise point set features directly instead electing to normalise the distance between two point set features during the similarity measurement process. That is, we normalised the difference between two point sets 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:

$$distance\left(q_{pts}, s_{pts}\right) = 1 - \frac{|q_{pts} \cap s_{pts}|}{|q_{pts} \cup s_{pts}|} \tag{6.7}$$

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

6.5 Retrieval Procedure

Algorithm 6.1 describes our overall retrieval process. We assumed that a database of graphs had been pre-constructed offline prior to the query process. We also assumed that the graphs in the index have been normalised. The query graph G_Q is also normalised (line 2). The normalisation operation uses the sets of feature means (F_{μ}) and standard deviations (F_{σ}) derived from the graphs in the index (see Appendix A).

The retrieval is then performed on the normalised query graph ${}^{N}G_{Q}$. The

retrieved images are sorted in ascending order of the graph edit distance between the query and indexed graphs (line 8).

6.6 Graph Comparisons

We compared graphs by adapting the beam search A^{*} algorithm [187], which calculates graph edit distances by applying a beam to the well-established A^{*} algorithm [186]. Both the standard brute force approach and its beam search adaptation assume a standard set of vertex features. Both the complete and CAPP graphs have different feature sets for different graph vertices (f_A for V_A and f_P for V_P). As such, it was necessary to adapt the beam search algorithm to account for these cases. The major challenges for this adaptation are:

- Maintaining the similarity of the graph structure, especially that of edges between vertices in V_A and V_P .
- Using different feature sets for different vertex sets.
- Avoiding substitution operations between vertices from different vertex sets so as not to compare tumours to anatomical structures.

Our graph comparison technique is presented by Algorithm 6.2.

Algorithm 6.1 Graph-Based Query Process		
1:	function QUERY(G_Q , index, F_μ , F_σ , beam)	
2:	${}^{N}G_{Q} \leftarrow \text{NORMALISEGRAPH}(G_{Q}, F_{\mu}, F_{\sigma})$	
3:	$similarities \leftarrow \emptyset$	
4:	for all ${}^{N}G_{S} \in index$ do	
5:	$sim \leftarrow \text{COMPARE}({}^{N}G_{Q}, {}^{N}G_{S}, beam)$	\triangleright sim is a 3-tuple $(img, dist, \varphi)$
6:	$similarities \leftarrow similarities \cup sim$	
7:	end for	
8:	$rankedList \leftarrow \text{SORT}(similarities)$	\triangleright sorts ascending using $dist$
9:	return rankedList	
10: end function		

Algorithm 6.2 Modality-Specific Beam-Search Algorithm

Require: $|G_Q| > 0, |G_S| > 0$ 1: function COMPARE $(G_Q, G_S, beam)$ $img \leftarrow I(G_S)$ 2: \triangleright The image represented by G_S $V_Q \leftarrow V(G_Q)$ \triangleright Obtain vertex set 3: $V_S \leftarrow V(G_S)$ 4: $P \leftarrow \emptyset$ \triangleright The search tree 5: for all $v_s \in V_S$ do 6: $\varphi \leftarrow \{\langle v_{q_1}, v_s \rangle\}$ 7: $\triangleright v_{q_1} \in V_Q, \varphi$ is a branch $dist \leftarrow \text{COST}(\varphi)$ 8: $P \leftarrow P \cup \{(dist, \varphi)\}$ 9: 10: end for $\varphi \leftarrow \{ \langle v_{q_1}, \emptyset \rangle \}$ \triangleright vertex deletion 11: $dist \leftarrow \text{COST}(\varphi)$ 12: $P \leftarrow P \cup \{(dist, \varphi)\}$ 13:loop 14: $P \leftarrow \text{TRIM}(P, beam)$ 15: $p_{min} = \underset{p \in P}{\operatorname{arg min}} dist\left(p\right)$ 16: $dist \leftarrow dist(p_{min})$ 17: $\varphi \leftarrow \varphi(p_{min})$ 18: $P \leftarrow P \setminus p_{min}$ 19: \triangleright removal of minimum cost match if COMPLETEMATCH(φ, V_Q, V_S) then 20: $\triangleright \varphi$ is completely expanded **return** $(imq, dist, \varphi)$ 21: else 22: $k \leftarrow |\varphi|$ 23:if $k < |V_Q|$ then 24:for all $v_s \in V_S \setminus \{v_{s1}, v_{s2}, ..., v_{sk}\}$ do $\triangleright v_{si}$ mapped in φ 25: $\varphi_{k+1} \leftarrow \varphi \cup \left\{ \left\langle v_{q_{k+1}}, v_s \right\rangle \right\}$ $dist_{k+1} \leftarrow \text{COST}(\varphi_{k+1})$ 26:27: $P \leftarrow P \cup \{(dist_{k+1}, \varphi_{k+1})\}$ 28:end for 29: $\varphi_{k+1} \leftarrow \varphi \cup \left\{ \left\langle v_{q_{k+1}}, \varnothing \right\rangle \right\}$ \triangleright delete vertex $v_{q_{k+1}}$ 30: $dist_{k+1} \leftarrow \text{COST}(\varphi_{k+1})$ 31: $P \leftarrow P \cup \{(dist_{k+1}, \varphi_{k+1})\}$ 32: else 33: $\varphi \leftarrow \varphi \cup \bigcup_{v_s \in V_S \setminus \{v_{s1}, v_{s2}, \dots, v_{sk}\}} \left\{ \langle \varnothing, v_s \rangle \right\}$ \triangleright vertex insertion 34: $dist \leftarrow \text{COST}(\varphi)$ 35: $P \leftarrow P \cup \{(dist, \varphi)\}$ 36: end if 37: end if 38: end loop 39: 40: end function

In our algorithm, set P (initialised on line 5) represents the graph edit distance search tree (described in Section 4.3.1 and shown in Figure 4.5). Each branch $p \in P$ is a 2-tuple consisting of a scalar value *dist*, the total graph edit distance of the branch, and a set φ , the search tree branch. The set φ contains mappings between vertices of the two graphs $\langle v_q, v_s \rangle$ where $v_q \in V_Q \cup \emptyset$ is a vertex of the query graph, $v_s \in V_S \cup \emptyset$ is a vertex of a graph in the index, and \emptyset is a non-existent vertex used to specify vertex insertion ($\langle \emptyset, v_s \rangle$) or deletion ($\langle v_q, \emptyset \rangle$). Each element of V_Q or V_S appears *at most* once in φ , while \emptyset may occur many times.

The search tree P (line 15) is reduced by removing branches ($p \in P$) until $|P| \leq beam$. The branches are removed in descending order beginning with those that have the highest *dist* values. Combined with the breadth-wise expansion of the search tree (see the loops beginning on lines 6 and 25), this means that the algorithm only explores those search branches that show the most promise in finding a minimal graph edit distance, i.e., they currently have the lowest *dist* values.

Every iteration of the main loop (lines 14 to 39), the branch with the minimum distance is selected from the search tree (line 16). The branch is a complete match if all elements of V_Q and V_S appear in φ once; if so, then φ is the optimal branch of the search tree and the function returns the image, its distance to the query, and the vertex mapping. This determination is made by the function COMPLETEMATCH(φ, V_Q, V_S).

If φ is not a complete match, then either there are still elements of V_Q or V_S or both that have not yet been mapped. In this case, if there are unmapped query vertices the algorithm expands the branch by mapping another vertex in V_Q (lines 25 to 32). Once all query vertices have been mapped, any remaining vertices in V_S are inserted (lines 34 to 36). The main loop then repeats until a complete match with a minimal cost is discovered.
Algorithm 6.3 Branch Cost Calculation						
1:	function $COST(\varphi)$					
2:	$dist \leftarrow 0$					
3:	$\varphi_t \leftarrow \emptyset$					
4:	for all $map_1 \in \varphi$ do	$\triangleright map_1 = \langle v_{q1}, v_{s1} \rangle$				
5:	$dist \leftarrow dist + Dist_p\left(\langle v_{q1}, v_{s1} \rangle\right)$	\triangleright Vertex distance				
6:	$\varphi_t \leftarrow \varphi_t \cup map_1$					
7:	for all $map_2 \in \varphi_t \operatorname{\mathbf{do}}$	$\triangleright map_2 = \langle v_{q2}, v_{s2} \rangle$				
8:	$\mathbf{if} \ map_1 \neq map_2 \ \mathbf{then}$					
9:	$dist \leftarrow dist + Dist_p\left(\langle v_{q1}v_{q2}, v_{s1}v_{s2}\rangle\right)$	\triangleright Edge distance				
10:	end if					
11:	end for					
12:	end for					
13:	return dist					
14:	end function					

6.6.1**Determining Operation Costs**

An important aspect of this algorithm is $COST(\varphi)$, the function that calculates the graph edit distance given a search tree path φ . This function (called $d(\cdot)$) in Equation 4.1) needs to account for several operations: substitution, insertion, and deletion. Furthermore, for each of these operations, the function must also calculate the cost of substituting, inserting, or deleting edges. The n-th vertex operation may result in up to n-1 edge operations, one for each vertex that was already in the branch prior to the current operation. This function is described by Algorithm 6.3.

Line 5 measures the cost to map two vertices. The cost for any edge operations is calculated by the inner loop (lines 7 to 11), which compares edges incident to $v_{q1} \in V_Q$ and any other vertex $v_{q2} \in V_Q$ to edges incident to $v_{s1} \in V_S$ and any other vertex $v_{s2} \in V_S$. The cost of the vertex mapping $\langle v_{q2}, v_{s2} \rangle \in \varphi$ has already been calculated in an earlier iteration of the outer loop. The edge operation costs are therefore calculated in an iterative manner.

The cost function used by the algorithm was given by the following equation:

$$Dist_{p}\left(\langle Q, S \rangle\right) = \begin{cases} 0 & \text{if } Q = \emptyset \text{ and } S = \emptyset \\ \infty & \text{if } mdt\left(Q\right) \neq mdt\left(S\right) \\ \left[\sum_{i}^{N-n} \left(q_{i}\right)^{p}\right]^{\frac{1}{p}} & \text{if } S = \emptyset \\ \left[\sum_{i}^{N-n} \left(s_{i}\right)^{p}\right]^{\frac{1}{p}} & \text{if } Q = \emptyset \\ \left[\sum_{i}^{N-n} \left(q_{i} - s_{i}\right)^{p} + \sum_{N-n}^{N} d_{J}\left(q_{i}, s_{i}\right)^{p}\right]^{\frac{1}{p}} & \text{otherwise} \end{cases}$$

$$(6.8)$$

where Q and S are a graph element (vertex or edge) of a query and stored graph, respectively; q_i and s_i are *i*-th feature of Q and S, respectively; p is the order of the equation; N is the total number of features in Q and S; n is the number of non-point set features in Q and S; and mdt(X) is a function that returns the modality of the ROI represented by a vertex (if X is a vertex) or the modality of the vertices incident to an edge (if X is an edge).

Equation 6.8 divides the cost function into five separate cases. Due to the design of Algorithm 6.3, the first case $(Dist_p = 0)$ can only occur on line 9 when the algorithm tries to calculate the cost to match two edges, $v_{q1}v_{q2}$ and $v_{s1}v_{s2}$, both of which do not exist. Since they both do not exist, no transformation operation is necessary, and as such the cost is 0.

The second case is a consequence of the multi-modality nature of our graphs. In Assumption 1 (see Section 5.1.1), we stated that anatomical data is provided by one modality, while pathology (or tumour) data is provided by the other. As such, when the modalities of the two elements, Q and S, are different, the cost is assigned to infinity to avoid matching vertices that represent anatomical structures with vertices representing tumours.

The final three cases were the costs for insertion, deletion, and substitution,

respectively. Our graph features consisted of three types (as described in Section 6.3); we therefore combined different functions to obtain our final cost. Measurement features and angle features were treated as elements of a feature vector. This allowed the costs to be determined by adapting the distance function D_p described in Equation 3.1. On the other hand, point set costs were determined using the Jaccard distance given by Equation 6.7. The insertion and deletion costs were equivalent to each other; in each case, the element that did not exist (\emptyset) was assumed to compose of numeric and angle features that were all normalised to 0. This was an adaptation of insertion and deletion cost calculations in prior work [183].

6.6.2 Beam Search Adaptation

The basic beam search A^{*} algorithm operates on a search tree created by breadthwise expansion. The tree is expanded iteratively by adding children to the current leaf nodes. During each iteration, the *i*-th vertex $v_{qi} \in V_Q$ is mapped to each vertex $v_s \in V_S$ that does *not* appear in the branch, as well as to \emptyset . Each of these mappings become new leaf nodes.

An example of such an expansion is shown in Figure 6.5. During the first iteration, the vertex A from the graph Q (Figure 6.5(a)) is mapped to all vertices of the graph S (Figure 6.5(b)), as well as to \emptyset , as shown in Figure 6.5(c). During the second iteration, each of the leaf nodes created during the first iteration is expanded. Figure 6.5(d) shows the expansion of the first leaf node ($\langle A, X \rangle$) of the tree. The vertex B from Q is mapped to all vertices of S (and to \emptyset) except for X, which has already been mapped in that branch. This expansion, which maps B to vertices of S, continues until all the leaf nodes created during the first iteration have been expanded. This is shown in Figure 6.5(e); the expansions of the nodes $\langle A, Y \rangle$ and $\langle A, Z \rangle$ have been left out for the purpose of clarity. Note that the expansion of $\langle A, \emptyset \rangle$ contains $\langle B, \emptyset \rangle$ as a child; \emptyset is allowed to appear





(c) Search tree after first iteration.



(d) Part of the second iteration; expanding the first leaf node.



(e) The search tree after the second iteration.

multiple times in any branch because it signifies the deletion of a vertex of Q.

Each leaf node represents a unique branch of the search tree (an element of P in Algorithm 6.2), and a unique combination of vertex mappings between a query and stored graph. The beam search algorithm reduces the search space at every iteration of the search tree expansion by expanding at most b leaf nodes. These

Figure 6.5: Search tree expansion for the graphs given by 6.5(a) and 6.5(b). During each iteration, all existing child search tree nodes 6.5(c) are expanded in turn 6.5(d). The new child nodes are not expanded until all previous child nodes have been expanded 6.5(d).

b leaf nodes represent those search tree branches that are most likely to deliver a low final total *dist*. Figure 6.6 shows the search tree for the graphs in Figure 6.5. This search tree was generated by the beam search algorithm with b = 2. Nodes that have been removed from consideration are marked in grey. During the first iteration, 4 nodes are added to the root. During the next iteration, only 2 of these nodes are expanded, for a total of 7 leaf nodes. In the third iteration, 2 of these 7 nodes are expanded, leading to the creation of 5 branches. This process continues until the final iteration where 3 nodes are reduced to 2; one of these two branches is the sequence with the best (lowest) beam search graph edit distance.



Figure 6.6: Search tree expansion with a *beam* value of 2. During each iteration 2 nodes are expanded a further level. The grey nodes indicate branches that are not expanded as they have been removed by the beam.

The multi-modality nature of our graphs and our formulation of the cost function given by Equation 6.8 opens the way for an additional optimisation of the search space. Any leaf node that contains a mapping of vertices representing ROIs in different modalities will have an infinite cost. We can therefore modify the beam search algorithm to not expand any branch that has an infinite cost. This further reduces the number of vertex mapping combinations that the A^{*} algorithm needs to consider in order to find the optimal graph edit distance cost. The function TRIM(P, beam) (line 15 of Algorithm 6.2) integrates this functionality alongside the normal beam search removal of leaf nodes in the search tree.

6.6.3 Aside: An Inferior Approach

The vertex set of our multi-modality graphs $V_K = V_A \cup V_P$ is by definition divided into two sets. Therefore, one could argue that a more efficient approach to the graph edit distance problem would be to apply the edit distance to each vertex subset independently and to combine the results. The motivation behind such an approach would be to reduce computational time. A brute force approach has exponential computational complexity based upon the order of the graph. By using two smaller vertex sets, one could reduce the time taken to calculate the graph edit distance. However, this approach is inferior.

Such a technique will mean that the graph edit distance for vertices in V_A will only map edges that are incident to two vertices in V_A . Similarly, the graph edit distance for vertices in V_P will only map edges that are incident to two vertices in V_P . Edges incident to vertices in V_A and vertices in V_P , the edges emphasised by the CAPP graph, are not mapped at all. These edges represent relationships between ROIs in different modalities and index features such as the distance separating a tumour from an organ. Cancer classification is dependent upon such features [171,172]; the independent calculation of graph edit distances on the vertex subsets is therefore not a viable option in our problem domain.

Chapter 7

Evaluation of Retrieval Method

This chapter contains the evaluation of our retrieval method. We evaluate both the accuracy and computational performance of our retrieval method.

7.1 Data Sets

We chose data sets to evaluate different aspects of our CBIR framework. A 2D liver shape data set was used to evaluate the retrieval precision on images with multiple shape distortions. We used a 3D simulated lymphoma data set to evaluate retrieval precision using images with multiple tumours placed across different anatomical structures. Finally, a PET-CT data set was used to investigate our method given the natural diversity of human anatomy and tumour distribution that exists in clinical practice. The clinical data set was also used to evaluate the retrieval of tumours in specific locations and for examining the contribution of different features. Table 7.1 summarises the data sets and how they were used in our experiments.

We compiled a reference index for all the data sets for use during our evaluation. Our definition of similarity considered two images to be relevant if the tumours within the images were located in a similar place and shared similar

Name	2D/3D	Size	Purpose
Liver Shapes	2D	500	- Evaluation of retrieval with multiple shape
			variations
			- Evaluation of computational performance
OncoPET	3D	50	- Evaluation of retrieval with multiple tu-
			mours spread across the whole body
PET-CT	3D	50	- Evaluation using clinical images (natural
			diversity of anatomy and tumours)
			- Evaluation of retrieval by specific tumour
			location
			- Evaluation of feature contribution
			- Evaluation of user interface (see Chapters 8
			and 9)

Table 7.1: Summary of Data Sets

characteristics (as stated in Section 5.1). This corresponds to the clinical staging criteria [171] where the classification of stage is dependent upon tumour properties (e.g., size) and location. We therefore created the reference indices by labelling every image according to the anatomical locations of their tumours and then listing all images with the same labels, as described in the following subsections.

7.1.1 Simulated 2D Liver Shapes

We used a set of simulated liver shapes to evaluate our approach on images with multiple shape distortions. The shapes were derived by applying multiple randomised variations to 50 manually segmented clinical MR slices¹. The 2D ROIs in every image were already delineated. To create multi-modality shapes for our experiments, we assumed that the anatomical ROIs (body, liver, and spine) were derived from MR and the tumour ROIs were derived from PET, i.e., we assumed the source clinical images were PET-MR. The data set contained several regions labelled as "unknown"; we marked these as tumour ROIs to increase the

¹The liver shape data set can be obtained from https://www.intelligence.tuc.gr/ ~petrakis/downloads/spatial-datasets-evaluation.zip.

diversity of PET shapes in our experiments.

We selected 500 images from the liver shape database. Images were divided into groups with the same number of tumours, and then sorted in ascending order according to liver size (area). We then selected images at equally spaced intervals to give us a consistent coverage of different liver shapes within the data set. Our selected subset included 100 images with 1 tumour, 100 images with 2 tumours, 100 images with 3 tumours, 100 images with 4 tumours, and 100 images with 5 tumours. The full data set has previously been used in other studies [82, 121, 183].

We created 6 reference indices for the liver shape data set: one for each set of 100 images with the same number of tumours, and one for the combined 500 image data set. During construction of the liver shape index we discovered that in addition to having tumours in the same relative location, every relevant image for a particular query also contained the same number of tumours as the query image, e.g., if a query had three tumours, every relevant image also had three tumours.

The feature set for the liver shape data was:

- $f_A = (s, b, l)$
- $f_P = (s, b, l, r)$
- $f_S = (d, ro, rs, md)$

Figure 7.1 shows several examples of the liver shape data. All the images in the data set contained 3 annotated anatomical regions: the liver, the spine, and the body outline. All other regions were marked as "tumour" or "unknown". In our experiments, we considered the "unknown" regions to be tumours, thereby increasing the diversity of tumour shape data.



Figure 7.1: Simulated 2D liver shapes. The arrows in the magnified image label the three known anatomical ROIs in every image. In the smaller images, the extra tumour and unknown regions are marked in red (left black in the magnified image).

7.1.2 OncoPET: Simulated 3D Lymphoma Images

The OncoPET database² [198] was created for the purpose of evaluating tumour detection algorithms and assessing computer-aided diagnosis methods. It contains 50 3D simulated ¹⁸F-FDG PET images, with multiple tumours per image. The tumours vary between 7mm and 14mm in diameter, have 5 different contrasts per diameter, and are spread across multiple sites: 180 in lymph nodes, 79 in the liver, 90 in the lungs and 26 in the spleen.

The OncoPET data set comes with two forms of annotation: a spreadsheet that defines tumour locations and a phantom image where different structures are labelled with different voxel values. The phantoms, originally used to create the simulated images, provided a perfect 3D delineation of the organs and tumours. We defined the ROIs as being representative of PET-CT images, assuming the anatomical ROIs were from CT and the tumour ROIs from PET.

We created a reference index for the OncoPET data using the tumour labels in the annotation. The OncoPET data set was used to evaluate the effectiveness of our algorithm on images containing multiple tumour in different locations.

²Available from: https://www.creatis.insa-lyon.fr/oncoPET_DB/



(a) Phantom

(b) Simulated image

Figure 7.2: An example of the OncoPET simulation data. The phantom (or model) for a lymphoma case and its associated simulated image (depicted with an inverted lookup table). The red arrows indicate the location of tumours in both the images.

The feature set for the OncoPET data was:

- $f_A = (s, b, l)$
- $f_P = (s, b, l, r, th, I_{MAX}, I_{\mu}, I_{\sigma})$
- $f_S = (d, ro, rs, md)$

Figure 7.2 shows an example of the images in the OncoPET data set. Figure 7.2(a) is the phantom image from which the simulated image (Figure 7.2(b)) was generated. Each grey-level pixel value in the phantom corresponds to a different tissue type. The red arrows indicate the tumours in the phantom and the corresponding tumours in the simulated PET image.

7.1.3 Clinical PET-CT Images

We collected 50 PET-CT studies of lung cancer patients on a Siemens Biograph mCT scanner. The reconstructed images 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 images in the data set contained between 1 to 7 tumours (inclusive). The studies included clinical reports detailing tumour locations and nodal involvement as specified by a clinician. The reference index for the clinical data set was derived from the tumour locations in the clinical reports. All data were de-identified.

We used a well-established adaptive thresholding algorithm [48] with refinements to segment the lung ROI from the CT. Tumours from the PET images were segmented with a 40% peak SUV connected thresholding to detect hot spots indicated in the diagnosed reports [199]. To include other major organs above the diaphragm we applied manual connected thresholding to coarsely segment the brain and mediastinal tissue (including the heart). We used the clinical reports to make minor corrections to the segmented ROI to ensure that the segments were well-defined.

The feature set for the clinical data was:

- $f_A = (s, b, l, ent, cont, cor, nrg, hmgt, pts)$
- $f_P = (s, b, l, r, th, I_{MAX}, I_{\mu}, I_{\sigma})$
- $f_S = (d, ro, rs, md)$

Figure 7.3 shows a sample of our clinical PET-CT data. Each PET-CT image pair (Figures 7.3(a) and 7.3(b)) was accompanied by a clinical report (Figure 7.3(c)). The clinical report contained information about the location of tumours in the image, written by an expert radiologist with several years of experience in PET-CT image interpretation.



(a) CT image

(b) PET image

Report

The patient was scanned with arms at the side. There are no large mass lesions in the cerebrum. There is good registration between the anatomical and functional data.

At the site of the <u>large mass</u> at the <u>posterior aspect of the right hilum</u> there is heterogeneous, probably partly necrotic, markedly glucose avid lesion which is consistent with a <u>high grade tumour</u>. Peak metabolism is 16.3 and the central

(c) Clinical report extract

Figure 7.3: An example PET-CT lung cancer image from our data set, and an extract from its associated clinical report. The images are shown in the coronal plane.

7.2 Experimental Procedure

7.2.1 Determining the Relevance of Retrieved Images

Our experiments were focused on determining if our method enabled the retrieval of multi-modality images. In addition, we investigated whether constraining the tumours to spatially related structures improved the retrieval accuracy; for this, we used the complete graph as a baseline that represented all features and relationships. We compared the retrieval results to the reference index for each data set. We evaluated our work using precision and recall, two standard metrics used to evaluate CBIR research (e.g., [25,82,116,121,183]). Precision is the proportion of retrieved images that are similar to the query, while recall is the proportion of similar images in the database that were retrieved. The ideal scenario would be 100% precision and 100% recall, implying that the CBIR method retrieved all the similar images and retrieved nothing but the similar images.

Precision and recall are defined as:

$$precision\left(k\right) = \frac{tp}{tp + fp} \tag{7.1}$$

and

$$recall\left(k\right) = \frac{tp}{tp + fn} \tag{7.2}$$

where k is the number of retrieved images, tp is the number of true positives (relevant retrieved images), fp is the number of false positives (not relevant retrieved images), and fn is the number of false negatives (relevant images that were not retrieved). Whether a retrieved image was tp, fp, or fn was determined by the reference index. For several experiments we also calculated the mean average precision (MAP), a single-value indicating the average precision over all levels of recall:

$$MAP = \frac{\sum_{k=1}^{Q} \left(\sum_{k=1}^{n} precision\left(k\right) \times \Delta recall\left(k\right)\right)}{Q}$$
(7.3)

where *n* is the number of images in the data set, *Q* is the total number of queries, and $\Delta recall(k) = recall(k) - recall(k-1)$, with recall(0) = 0.

Our retrieval experiments used a leave-one-out (LOO) cross-validation approach. We divided every data set of m images into m sets consisting of 1 query image and m - 1 indexed images. The removal of the query from the indexed image ensured that our final precision and recall were not biased. We performed

LOO retrievals on all data sets using both complete and CAPP graphs and compared them by calculating the mean precision over all values of recall. For our clinical data set, we also compared the complete and CAPP graphs based on the location of tumours. We also manually examined the visual similarities of the retrieved images and the query.

We also examined the contribution of the image features on PET-CT retrieval using CAPP graphs. 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 $1 \leq f \leq |F|$. The MAP of LOO retrievals using each F_i was then calculated. We also examined the contribution of the graph structure by repeating the LOO retrievals using no image features.

7.2.2 Evaluating the Computational Performance

The computational evaluation was performed on an Intel Core i5 CPU clocked at 2.67 GHz with 4 GB RAM, running Windows 7 Professional 64-bit. Our graph comparison algorithm was implemented in MATLAB 7.11 (R2010b) 64-bit.

The computational performance was evaluated using the 5 liver shape subsets divided based upon the number of tumours in each number (Section 7.1.1). We measured the time taken by our implementation of Algorithm 6.2 when comparing two CAPP graphs representing images with the same number of tumours. This calculation was repeated in a LOO approach for all images with the same number of tumours, thereby producing 9900 graph comparison times for each of the 5 subsets. We repeated this procedure for 10 different beam sizes, ranging from 100 to 1000, increasing in increments of 100.

Our experiments allowed us to investigate and compare the time taken by our algorithm when using graphs with different orders. Furthermore, by varying the beam we were also able to examine the impact of beam size on the time taken to compare graphs.



Figure 7.4: Liver shape retrieval (by graph order): mean precision vs. recall. This figure shows the mean precision and recall of liver shape retrieval with complete and CAPP graph representations on a data set containing images with the same number of tumours.

7.3 Results

7.3.1 Liver Shapes

We examined the retrieval performance of the two graph representations with a typical scenario of retrieving images with the same number of tumours and localisations as the query. Figure 7.4 shows the mean precision of these experiments. The complete and CAPP graph representations were able to retrieve the images from the data set. Figure 7.4 also shows that the CAPP graph achieved a higher average precision for every subset.

We also compared the retrieval of the liver shapes using a data set containing images with varying numbers of tumours (combination of all the images from the previous experiment). Figure 7.5 shows the precision and recall of the two



Figure 7.5: Liver shape retrieval (by graph order): mean precision vs. recall. This figure shows the mean precision and recall of liver shape retrieval with complete and CAPP graph representations on the data set of 500 images with different order number of tumours.



Figure 7.6: Example of liver shape retrieval. Anatomical regions are bordered in black while tumour ROIs are marked in red. Note that we considered unknown ROIs to be tumours to include greater shape diversity. The graph edit distance is given beneath each retrieved image. The complete and CAPP graph edit distances are independent of each other, and serve as a measure of the difference between the retrieved image and the query for the purpose of retrieval ranking.

representations.

An example of liver shape retrieval on the 500 image data set is shown in Figure 7.6. The query image contains 5 tumours: one in the liver, and 4 in the body (2 of which are overlapping) depicting a typical scenario of retrieving images with tumours in different anatomical locations. The graph edit distance is provided for each retrieved image; the distances for the complete and CAPP graph methods are independent of each other.

7.3.2 OncoPET

Figure 7.7 depicts the mean precision and recall of our retrievals on the OncoPET data set. A retrieval example with the OncoPET data set is depicted in Figure 7.8; the images are given as maximum intensity projections (MIPs), which are 2D reconstructions of 3D data that show the pixels with the highest intensity from the given point of view. Table 7.2 lists the localisation of the five query tumours and the tumours of the top three retrieved images using the complete and CAPP graphs; this is done to clarify the location of the tumours in Figure 7.8.



Figure 7.7: OncoPET retrieval: mean precision vs. recall.



Figure 7.8: Example OncoPET retrieval. Note that in these images the right side of the body appears on the left. The figure shows maximum intensity projections (MIPs) of the simulated PET images with tumours marked in red. Since there is a loss of depth in MIPs, the exact tumour locations are provided in Table 7.2. The edit distances from the query graph are given beneath the retrieved images, which have been cropped to the chest region for clarity.

		Tumour Localisation ^a					
		1	2	3	4	5	
Query		LL	LL	RL	HT	LL/HT	
	R1 ^b	LL	LL	RL	HT	HT	
Complete	R2	LL	LL	RL	LL/HT	LL/HT	
	R3	LL	LL	LL	LL	$\mathrm{RL/HT}$	
	R1	LL	LL	RL	HT	LL/HT	
CAPP	R2	LL	LL	RL	HT	HT	
	R3	$\mathbf{L}\mathbf{L}$	LL	RL	LL/HT	LL/HT	

Table 7.2: OncoPET Retrieval Localisation

^a LL = left lung, RL = right lung, HT = heart, X/Y = near X and Y. ^b RN = *n*-th retrieved image.

7.3.3 Clinical PET-CT

Figure 7.9 depicts the mean precision and recall of the retrievals carried out on the clinical PET-CT data set. Table 7.3 provides a breakdown of the retrieval accuracy (using the MAP) by tumour location. Table 7.4 lists the effect of the image features on the retrieval process; we calculated the MAP of CAPP graph retrievals using all features, using no features, and using our varied feature set



Figure 7.9: PET-CT retrieval: mean precision vs. recall.

(excluding one feature at a time).

A PET-CT retrieval example is depicted in Figure 7.10. Table 7.5 gives the the locations of tumours in query and the top three retrieved results using both the complete and CAPP graphs.

7.3.4 Computational Performance

Figure 7.11 shows bar charts that depict the mean retrieval time based on the number of tumours in the images. The bar groups show the effect of the different beam sizes for each set of images with the same number of tumours. Due to issues of scale, we have separated the results for images with 1 and 2 tumours (Figure 7.11(a)) from the results for images with 4 and 5 tumours (Figure 7.11(b)). The results for images with 3 tumours are replicated in both these charts to provide a reference for the relative time difference. The bar heights are the mean

Location	MAP (%)		
Location	Complete	CAPP	
right upper lobe (RUL)	46.06	50.27	
right middle lobe (RML)	12.51	12.20	
right lower lobe (RLL)	24.51	24.18	
left upper lobe (LUL)	14.45	23.17	
left lower lobe (LLL)	25.06	26.78	
mediastinum (M)	36.64	39.14	
right mediastinum (RM)	14.00	24.68	
right hilum (RH)	37.32	45.52	
left hilum (LH)	30.14	45.30	

Table 7.3: MAP By Tumour Location

 Table 7.4:
 Feature Contributions

		MAP (%)			
CAPP - all features used		52.28			
complete - all	features used	46.58			
CAPP - no	features used	46.12			
Vertex Feature	MAP (%)	Edge Feature	MAP (%)		
size (s)	52.44	distance (d)	52.68		
boundary (b)	52.50	relative size (rs)	53.40		
length (l)	52.47	relative orientation (f)	48.40		
entropy (ent)	52.34	minimum distance (md)	52.20		
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				



Figure 7.10: Example PET-CT retrieval showing MIPs of the PET images, with tumour ROIs marked in red. The edit distances from the query graph are given beneath the retrieved images, which have been cropped to the lung fields.

		Localisation Labels ^a					
		1	2	3			
Query		RLL	RH	-			
	R1 ^b	RUL	RH	М			
Complete	R2	LUL	LH	М			
	R3	RUL	RH	Μ			
	R1	М	RH	-			
CAPP	R2	RUL	RH	Μ			
	R3	RLL	RH	-			

Table 7.5: Clinical Retrieval Localisation

^a RLL = right lower lobe, RH = right hilum, RUL = right upper lobe, LUL = left upper lobe, LH = left hilum, M = mediastinum.

^b RN = n-th retrieved image.

time while the error bars depict 1 positive standard deviation from this mean time.

In a similar manner, Figure 7.12 shows bar charts depicting the mean retrieval time based on varying beam sizes, with the bar groups indicating the effect of an increasing number of tumours in the images. Once again, due to issues of scale, we have separated the bar groups for images with 1 and 2 tumours (Figure 7.12(a))



(a) Graph comparison time by number of tumours (images with 1 to 3 tumours)



(b) Graph comparison time by number of tumours (images with 3 to 5 tumours)Figure 7.11: Graph comparison time by number of tumours.



(a) Graph comparison time by beam size (images with 1 to 3 tumours)



(b) Retrieval time by beam size (images with 3 to 5 tumours)

Figure 7.12: Graph comparison time by beam size.

Boom	Measurement (seconds)							
Dealli	mean	\min	max	25%	50%	75%	95%	
100	0.3435	0.0512	0.7382	0.2564	0.3248	0.4287	0.5654	
200	0.6490	0.0530	1.8332	0.4216	0.6032	0.8301	1.2854	
300	0.9810	0.0519	3.2439	0.5288	0.8712	1.2825	2.2412	
400	1.3502	0.0531	5.1745	0.6245	1.1402	1.8019	3.3938	
500	1.7381	0.0521	8.4201	0.7198	1.3651	2.3162	4.7461	
600	2.1437	0.0531	12.4875	0.8129	1.5849	2.8573	6.1160	
700	2.5695	0.0521	17.4058	0.9024	1.8151	3.4413	7.5317	
800	3.0036	0.0525	22.1322	0.9861	2.0488	3.9873	8.8613	
900	3.4326	0.0524	27.7739	1.0590	2.2852	4.5279	10.0562	
1000	3.8516	0.0524	33.9689	1.1167	2.5252	5.0970	11.5476	

Table 7.6: Graph Comparison Time Summary (Images with 5 Tumours)

from the bar groups for images with 4 and 5 tumours (Figure 7.12(b)). The results for images with 3 tumours are replicated in both these charts to provide a reference for the relative time difference.

Table 7.6 and Figure 7.13 provide further analysis of the graph comparison time in the case of graphs representing images with 5 tumours (the largest graphs for the liver shape data set). The table shows for different beam values the mean, minimum, and maximum graph comparison times. It also shows the retrieval times at different percentiles: 25% (1st quartile), 50% (median), 75% (third quartile), and 95%. The figure shows histograms of the graph comparison times using the liver shape data set with 5 tumours per image; each histogram was derived using a different beam value. The histograms contain 100 bins each, with the range each bin representing 1% of the maximum time.

7.4 Discussion

During the liver shape retrievals we discovered that the CAPP graph had a higher retrieval precision than the complete graphs for recall < 40% (see Figure 7.5).



Figure 7.13: Histograms of graph comparison time using the liver shape data set with 5 tumours per image. Each histogram has 100 bins, scaled to the maximum retrieval time.

The complete graph had consistent retrieval precision for levels of recall $\geq 30\%$ and had a higher precision than the CAPP graph at levels of recall $\geq 40\%$; the CAPP graph's retrieval precision continued to degrade linearly. This was due to the characteristics of the data set. According to the liver shape reference index, all similar images contained the same number of tumours as their associated query and therefore had graphs with the same order. Complete graphs inherently favoured this situation because the major distinction between complete graphs is order rather than structure. Further analysis of this is given by Figure 7.14, which plots the average difference in the number of vertices in the query and retrieved graphs against the similarity rank of the retrieved images.



Figure 7.14: Mean difference in retrieved graph order; the mean difference is plotted against rank of the retrieved image.

In the figure, the complete graph plot follows a step function with a jump at approximately every 100 retrieved images. The flat gradient of the plot in the first 100 images at a mean difference of 0 means that the complete graph method guaranteed that most of the similar images would be retrieved within the first 100 images. This was only possible because in the data set every similar image contained the same number of tumours as the query. On the other hand, the CAPP approach retrieved graphs of different orders as evidenced by the erratic nature of its plot and thus could not guarantee that all of the similar images would be retrieved within the first 100 images.

These results indicated that the CAPP graph found the most similar liver shapes earlier than the complete graph. In retrieval applications, the most similar results are expected within the first few retrieved images as evidenced by the use of precision at 4 or 5 used by Shyu et al. [100] and Quellec et al. [116]; the CAPP graph is more capable than the complete graph at meeting this criterion. Furthermore, Figure 7.4 shows that the CAPP graph would maintain a higher retrieval precision than the complete graph if the data set was filtered to only consider images with the same number of tumours as the query.

The precision and recall in the OncoPET retrievals demonstrated that the CAPP graph had a higher precision than the complete graph when retrieving images with multiple lesions placed across different organs in a 3D space. It also demonstrated the accuracy of our method when the images contained a fixed anatomy containing a large number of tumours. In these experiments, the CAPP graph achieved a maximum precision that was approximately 40% higher than that of the complete graph. The OncoPET images all contained a large number of tumours (5 or 10) and the results demonstrated that in such complex cases the CAPP graph had precision $\geq 60\%$ at levels of recall $\leq 50\%$. Every query had very few similar images (average 2.2, deviation 1.5) as defined in our reference index (Section 7.1.2) and as such every false positive resulted in a large precision drop; this impacts on the complete graph approach more than the CAPP graph due to the lower discriminatory power of the former.

Our clinical retrievals demonstrated that the CAPP graph had a higher overall mean precision than the complete graph (Figure 7.9). The CAPP graph also had higher a MAP when retrieving tumours in most anatomical locations (Table 7.3), with right middle lobe (RML) and right lower lobe (RLL) tumours being the exceptions. However, in both the exceptions the difference in MAP was < 0.4%. There were also very few RML and RLL tumours in the data set (5 and 9, respectively) and the images often contained other tumours. As such, the retrieved images often had other similarities to the query (e.g., anatomical similarities, other tumours in similar locations) but did not have a tumour in the particular location of interest. Weighting the retrieval process can overcome this hurdle. This is a non-trivial task that requires weighting not only individual features but also a subset of the query graph structure; this is a non-trivial task and is left for future work.

Our retrieval examples (Figures 7.6, 7.8, and 7.10) also demonstrated that the images retrieved by the CAPP graph were visually more similar (based on relative tumour location) than those retrieved by the complete graph. Visual inspection of Figure 7.6 shows that the images retrieved by the CAPP graph always contained 1 tumour in the liver and 4 in the body, exactly the same as the query. On the other hand, the third image retrieved by the complete graph method has two tumours within the liver, while the fourth retrieved image has three. The discriminatory power of the CAPP graph is apparent by comparing the graph edit distances across the retrieved ranks. Retrieval using both representations was able to find the exact match as the first retrieved image. For the complete graph approach, the increases in the graph edit distance between the second and third, third and fourth, and fourth and fifth retrieved images were 9.3%, 3.9%, and 0.2%, respectively. Similarly, the increases in the graph edit distance for the CAPP graph approach were 21.5%, 7.3%, and 1.3% respectively. The larger differences in graph edit distances also indicate that CAPP graphs are easier to discriminate when ranking image similarity.

Similarly, Figure 7.8 shows that the OncoPET images retrieved by the CAPP graphs all contained tumours localised in the same anatomical region as the query. The 3rd image retrieved by the complete graph method did not have a similar localisation, instead containing 4 tumours within the left lung and 1 tumour located in the region between the right lung and the heart (see Table 7.2).

In the clinical retrieval example (Figure 7.10), only the CAPP graph was able to retrieve an image that shared all tumour locations with the query. The second image retrieved using the complete graph did not have any tumours localised near the same structure as in the query. All other retrieved images contained at least one similar localisation: a tumour that affects the right hilum. The clinical results demonstrated that our CAPP graph approach is capable of retrieving similar PET-CT images with the realistic, natural variation of in anatomy and tumour location that occurs in human patients. However, weighting of particular substructures is necessary to emphasise the tumours in particular locations of interest.

Our findings revealed that the CAPP graph had higher retrieval precision than the complete graph in most retrieval scenarios. The higher precision of the CAPP graph over the complete graph across all three data sets 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 Figure 7.15. 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.

Equation 6.1 created CAPP graphs by preserving edges between tumour vertices and the vertex of its nearest anatomical neighbour. Modifying Equation 6.1 to include more edges (such as by preserving edges with the two nearest anatomical neighbours) would cause a decrease in retrieval precision. In fact, increasing the number of edges (to the n closest organs) moves the CAPP graph's structure closer to that of a complete graph. We carried out an experiment to test this possibility. CAPP graphs that preserved edges with the two nearest anatomical neighbours achieved a MAP of 50.51% compared to MAPs of 46.58% using the complete graph and 52.28% using the CAPP graph as defined.



Figure 7.15: Comparison of graph structures of the same order but having more than one tumour. Complete graphs of the same order always have the same structure, while CAPP graphs may have more varied structures. Column (a): The image depicting four anatomical regions (grey ROI) and two tumours (white ROI). Column (b): The complete graph structure that represents all the images. Column (c): Distinct CAPP graph representations.

In addition, our similarity measurement algorithm extended one that emphasised inter-class distances [187]. Calculating the similarity of these images with complete graphs tends to favour the use of substitution operations, especially when the images have the same number of tumours. When CAPP graphs are used, all operations are applied. As defined, insertion and deletion operations have a higher cost (see Equation 6.8). This results in CAPP graphs having greater discriminatory power. As such, the CAPP graph had the highest precision when the graph data set contained a wide variety of different structures, representing images with many tumours spread across the body. This is illustrated by the levels of precision achieved when the CAPP graph was used to retrieve OncoPET images (Figure 7.7) and clinical images (in Figure 7.9).

Our choice of graph representations was primarily motivated by their influence on retrieval based on the spatial arrangement of the image content [121]. In our graph representation each vertex or edge was treated as a feature vector. Thus the image features (see Section 6.3) and the distance metric (Equation 6.8) played a critical role during similarity calculation by allowing the graph matching algorithm to measure the level of similarity between these vectors. A small difference between two feature vectors indicated a low cost graph edit operation (implying a minor graph transformation). Table 7.4 shows that retrieval based purely on graph structure had a lower MAP than retrieval that considered both image features and graph structure. From this we can conclude that the image features contributed to the precision of our retrieval method and the retrieval precision of the CAPP graph was not purely the effect of the graph structure.

Table 7.4 also shows that the CAPP graph was able to maintain similar levels of retrieval precision using different feature sets. Our normalisation scheme (see Section 6.4) ensured that the maximum and minimum possible contribution of each feature to the distance metric (Equation 6.8) was the same. This meant 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 lung lobes. 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. Image similarity based on disease localisation requires that anatomical structures be correctly matched across images. Normally, this can be done by matching labels assigned to the segmented anatomical ROIs. However, to allow segmentation that was not specific to a structure we assumed in Section 5.1.1 that the segmented structures were not labelled. Therefore, in our case the anatomy assignment was entirely dependent upon the graph similarity calculation. An incorrect anatomy mapping would potentially lead to an incorrect tumour mapping, e.g., an incorrect match between a liver and a lung would cause lung tumours to be considered as liver tumours. 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 complete anatomical subgraph. This complete anatomical subgraph did not hinder the retrieval because most humans bodies have a similar spatial arrangement of organs; thus it was more important to represent all spatial relationships.

Our performance evaluation revealed that the time taken to compare two graphs increased exponentially as the number of tumours represented by the images increased (a larger number of tumours results in more graph vertices, i.e., a higher graph order). This increase was expected. The beam search algorithm is still a brute force approach; its restriction of the search space simply makes it find a solution faster than a pure brute force approach that considers all possibilities [187]. This is also indicated in Figure 7.12; the mean comparison time increases with the number of tumours for any fixed beam size.

Figure 7.11 indicates a linear relationship between retrieval time and beam size for images with the same number of tumours. This is clearly seen in Figure 7.11(b) for the case where there are 5 tumours. When beam = 1000, the mean time is under 4 seconds; this time is reduced to less than 1 second when the beam is reduced to 100, i.e. the reduced search space of a smaller beam reduces the time taken by the algorithm.

According to the detailed comparison of the graph matching times (Table 7.6 and Figure 7.13), the minimum times are almost exactly the same regardless of the beam size (standard deviation of about 0.0006s). This indicates consistency in calculating the graph edit distance to the most similar structures where minimal search tree expansion is required, such as cases without vertex deletion or insertion. Furthermore, comparing the percentile times to the maximum times indicates that the maximum is not representative of the other comparison times. In our experiments, during the worst case (beam = 1000) three-quarters of the data set took one-sixth of the time of the worst case. A further 20% of the data set took about one-third the time of the maximum. Figure 7.13 clearly shows that the distribution is heavily weighted to the left (lower comparison times). As such, while the worst case computational performance of our retrieval algorithm is not ideal, we have shown that the majority of comparison times are relatively feasible.

7.5 Summary

The CAPP graph representation enabled the retrieval of both 2D and 3D multimodality medical images based on the localisation of tumours in relation to anatomy. Our results demonstrated that the varied nature of CAPP graph structures enabled higher levels of retrieval precision compared to complete graphs, which have no structural distinction. These capabilities arise from the CAPP graph's ability to model disease localisation within multi-modality images. The CAPP graph and our proposed similarity measurement algorithm also allowed the use of modality-specific features to capture the complementary information inherent in multi-modality images. This validates our hypothesis that the retrieval of multi-modality images can be improved by emphasising tumour localisation through a graph representation that strongly associates individual tumours and their nearest organs.

Our computational evaluation indicated that the graph comparison time was still exponential according to the number of graph vertices. Furthermore, we discovered a linear relationship between the beam size and comparison time for graphs with the same number of vertices. These results suggest that there needs to be further study into the balance between the graph order, beam size, and comparison time. In particular, we need to investigate a method for deciding upon the most appropriate beam size for any given graph order that allows retrieval to be carried out in real-time.

Chapter 8

Graph-Based Retrieval Interpretation

In this chapter, we describe a method that assists users in interpreting multimodality images that have been retrieved by our CBIR framework. In particular, we propose recommendations for the design of CBIR user interfaces (UIs) and a method for assisting user interpretation of retrieved three-dimensional and multimodality medical images.

8.1 Integration with the CBIR Framework

Figure 5.1 shows the way in which our UI for interpreting the retrieved results integrates with our retrieval engine, which was presented in Chapter 6. The UI takes the similarities calculated by the retrieval engine and uses it to display the retrieved images in a ranked order. The visualisation of the retrieved images integrates this ranking along with individual graph representations, clinical reports, and the ROIs for every image. The interactions defined by the UI, e.g., search filters, will allow users to explore and understand the set of retrieved images.
8.2 UI and Visualisation Requirements

The presentation of the retrieved results is of paramount importance as it allows the user to interpret the images for whatever purpose they executed the search. Thus it is necessary that the visualisation of the retrieved results enhances data analysis and assists a user in organising their mental interpretation of the presented data.

Tory et al. [96] presented several guidelines for domain or task specific visualisations, such as grouping of related information, imposing structure on data and tasks, and abstracting and aggregating material. In addition, Wilson [95] made several recommendations for the design of retrieval UIs, which included guidelines for the presentation of the query, retrieved data, and metadata, as well as the level of control available to users. For application to images such as PET-CT, the guidelines suggested multiple visualisations for different types of data (anatomical, functional, fusion) as a means of assimilating the different views provided by PET-CT. The use of abstractions and supplementary data would allow complex PET-CT images to be summarised.

Based on the characteristics of PET-CT, we combined the visualisation [96] and search UI [95] guidelines to elicit the following requirements for enabling effective user interpretation of retrieved images:

- The user interface must be capable of displaying volumetric PET-CT images. Multiple images should be visualised simultaneously to allow for visual comparison on the same screen. This includes the simultaneous visualisation of the query and retrieved images.
- 2. The user interface should show an abstraction of the PET-CT images.
- 3. The interface should allow the user to browse through all retrieved results, not necessarily the k most similar images (see the k-nearest-neighbour

search scheme briefly described in Section 3.2 on page 27).

4. The interface should display related patient data, e.g., clinical reports.

We also elicited several UI interactions that could potentially enable better understanding of the retrieved images by users:

- The user should be able to switch between different views of the same image,
 e.g., fused and non-fused PET-CT or 3D projections.
- 6. The user interface should allow the user to change the presentation of the data, e.g., reordering, grouping, filtering, or searching within the retrieved data.
- 7. The interface should allow the user to provide feedback by adapting or refining the query.
- 8. The user should be able to interact with the abstraction of the image in order to enhance their understanding of the PET-CT images. This could be through linking the abstraction to important parts of the supplementary data (patient reports) or by highlighting important aspects of the images, similar to the vertebrae outlining in [110].

8.3 User Interface Design

Figure 8.1 shows the general layout of our UI. We designed the interface with five main partitions. The query partition displays the query image and its abstraction; similarly, the section for viewing retrieved images displays multiple retrieved images and their abstractions. We display any available supplementary information for a selected image (whether query or retrieved) in the section set aside for this purpose. A set of navigational components for browsing through all the retrieved images is provided in the lower right corner, while functions



Figure 8.1: Retrieval UI design schematic.

for searching, sorting, and filtering and grouping the results are provided in the lower left corner. Three of the five divisions were absolutely necessary: the query partition, the retrieved image partition, and the navigational tools. The function bar was added to allow interactions for further exploring the retrieved image data set, while the supplementary image panel was added to allow access to clinical reports and other non-image information without the need to access an external application (e.g., a PACS system).

The area for displaying the retrieved images is divided into four partitions. Each partition shows a different retrieved image; the most similar retrieved image is placed in partition 1, the second most similar image in partition 2, and so on. When the user navigates to a different page of retrieved results, then partition kcontains images with retrieved rank equal to $k + 4 \times (page - 1)$.

The implementation of our design is shown in Figure 8.2. This UI layout inherently fulfills several of the visualisation requirements: 3, 4, and the multiple image visualisation part of Requirement 1. The query and retrieved images are



Figure 8.2: Retrieval UI implementation

displayed simultaneously on the same screen. The navigational controls allow the user to browse through the entire retrieved image set and not just the first 4 retrieved results. Finally the area for displaying supplementary information allows the system to present any extra, related information, e.g., by displaying clinical reports.

8.3.1 Visualising Multiple Views of Volumetric Images

Individual PET-CT images require relatively high storage space compared to regular images (several hundred megabytes instead of several kilobytes for a JPEG). As such, displaying multiple PET-CT is a memory intensive task even before we consider displaying multiple views of the same image. Long load times are also expected for such images.

We have undertaken two strategies to feasibly display multiple views of volumetric PET-CT images (or any other image with similar memory requirements). First, the preprocessing performed by our framework (Figure 5.1) converts the DICOM images to fast-loading TIFF stacks that are lossless compressed and downscaled (reduced pixel resolution), thus reducing the memory footprint of the individual images [196]. The preprocessing phase also calculates views for navigating the images in advance instead of at run-time. These views included the coronal CT and PET stacks, a fused PET-CT stack with a colour lookup table, and MIPs of the PET images. The coronal views were further downscaled to ensure that each slice could be completely visualised within the UI.

The second strategy is in a view-on-demand approach for image visualisation. Rather than displaying all views of all images currently on the screen, we show specific views only when requested by the user. The unseen views are loaded into memory in preparation for display. This allows us to switch between views almost instantly; this is an important capability for understanding multi-modality images where users need to integrate information from different images. When the user navigates to a new page, these images are unloaded from memory and are replaced by the images on the new page. Keeping unseen images outside of primary memory gives our CBIR UI a consistent memory footprint. At the same time, loading new images does not take a long time because each of the views has been pre-calculated and stored as a fast-loading TIFF stack.

Figure 8.3 shows the way in which we display different views of the same data. Each image view is loaded into a separate tab, labeled with the name of the view. When the tab is selected, the image stack corresponding to the view is displayed. This display method conserves screen space but still allows the user to assimilate information from multiple views. This UI component fulfils Requirement 5 and



the volumetric image visualisation part of Requirement 1.

Figure 8.3: Switching between views of the same data. When the user selects a tab, the currently displayed image stack is replaced with the new image. (a) is the coronal CT view of a PET-CT scan, (b) is the fusion of the coronal PET and CT scans, rendered with a default fusion ratio, and (c) is the MIP generated from the PET image.

8.3.2 Abstraction Visualisation

Abstractions of complex data allow humans to gain an overall understanding of the visualised information without the need to examine it in great detail [96]. In the case of PET-CT CBIR, an abstraction can inform users about the overall properties of the image before they decide to interpret it in detail. This is of particular interest in cases that could be time consuming to interpret, e.g., images with multiple tumours interacting with multiple organs or where there are a large number of cases to be viewed.

In our framework, the CAPP graph representations of the PET-CT images form a natural abstraction of the information encoded in the two related volumes: the graph vertices abstract the ROIs, while the graph edges encode spatial relationships between these ROIs (see Chapter 6). Furthermore, we use the features to affect the layout and properties of the visualisation in two ways: (i) the size of each vertex is proportional to the volume of the 3D ROI it represents, and



Figure 8.4: Visualising a graph abstraction and its corresponding PET-CT image. Each graph vertex represents an ROI within the image, as explained by the legend (b). Vertex positions are determined by the relative position of the ROIs within the actual image; vertex sizes are similarly derived from the ROI volumes.

(ii) positioning tumour vertices in the visualisation according to the relative position of the tumour ROI and anatomy ROI in the PET-CT images. The graph abstraction visualisation fulfilled Requirement 2.

As such, our graph visualisation enabled users to interpret images based on the location of tumours in relation to anatomy without needing to physically navigate through the images to find the tumours. For example, when searching for a tumour near a particular location the graph visualisation enabled the user to decide whether the retrieved image was relevant to his or her search criteria. An example of our graph visualisation is depicted in Figure 8.4.

8.4 User Interactions

The interactions provided by a UI are an essential tool for two-way communication between the system and the human using it. We defined several mouse-based interactions for our CBIR UI, designed to allow users to seek out the information that would assist them in interpreting the retrieved images in the context of their query.

8.4.1 Changing Data Presentation

Customising the way retrieved data is presented allows users to more quickly locate the images they require. We fulfilled Requirement 6 by providing several methods for the user to select the way they accessed retrieved data.

Sorting Options

We provided two options for sorting the retrieved data: by image similarity (as calculated by the CBIR algorithm), or by the complexity of the pattern, i.e., the number of ROIs (corresponding to the number of graph vertices). In the second case, retrieved images with the same complexity were sorted according to image similarity (from the most to the least similar).

Search by Features

We implemented tools that allowed the user to search within retrieved results, based upon the features of individual ROIs. For example, the user could specify a range of values for the volume of a tumour; this would result in the visualisation of only those retrieved images that contain a tumour whose volume falls within the specified range.

Filtering and Grouping

We allowed the user to filter or group the retrieved results based on the proximity of tumours to other structures. This functionality was implemented by analysing the graph representations of the images; the graph edge features indicated whether tumours were located in or near an organ, either through total or partial inclusion. Each filter was defined as a pairing of tumours with an anatomical structure, i.e., "tumours + organ". Multiple filters could be applied to further constrain the images that were displayed; for example, when the user applied filters for "tumours + left lung" and "tumours + right lung" only those similar images with tumours in both lungs were displayed, and other cases (such as images with tumours only in the mediastinum) were removed from the results that were displayed.

8.4.2 Query Refinement

Our method of query refinement (to fulfil Requirement 7) allowed the user to edit aspects of the graph representation through two abstraction-based interactions. The first interaction allowed the structure of the graph to be edited, e.g., the removal of vertices (ROIs) or edges (relationships). The second interaction applied weights to individual features of individual vertices to modify their importance to the query. For example, the user could increase the weight applied to a tumour's volume and leave all other tumour features the same, while also decreasing the weight of the left lung's volume and increasing the weight of all left lung features.

Examples of the query refinement interactions are shown in Figure 8.5. In Figure 8.5(a), the structure of the query graph has been changed; the smaller tumour vertex has been deleted, as have the edges between the mediastinum and the brain, and the mediastinum and the remaining tumour. In Figure 8.5(b), an increased weight has been applied to the *volume* feature of the selected tumour. The graph abstraction has been updated to reflect this change visually: the vertex size has changed. These refined graphs can now be used as updated queries for the CBIR engine.

8.4.3 Visual Indication of Similarity

We presented three methods by which users could interpret image similarity. The first method was visual inspection of multiple image views and displaying the similarity (or dissimilarity) value calculated by the CBIR engine. The second



(b) Editing the features of the query graph.

Figure 8.5: Query refinement.

method was the standard approach of giving users access to the clinical reports. Users were then able to view the images and read the reports to find similarities between the query and retrieved cases, both visually and in the text. However, this approach did not assist the users in visually finding similarities among images.



Figure 8.6: Understanding PET-CT similarity through visually-driven user interactions. The user selects a vertex from a graph visualisation. The vertex mappings calculated by the CBIR engine during retrieval are used to mark the most similar vertices on the other graphs. This information is combined with the database of segmented ROIs to highlight the most similar regions on the images. The initial selection is not limited to the query graph.

Our final method used the abstractions (graph visualisations) as a way of explaining the similarity between the query and the retrieved images (Requirement 8). More specifically, we defined an interaction for every visualised graph vertex: upon user selection, a vertex would gain a unique border to indicate that it was selected. At the same time, the ROI corresponding to the selected vertex would be outlined within the image. In addition, the image features indexed on the selected vertex would be displayed to the user. The most similar vertices in all other displayed graphs would also be given a unique border and their corresponding ROIs would also be highlighted in their associated images. The overall aim of this interaction was to allow the user to immediately find similar ROIs across different images. An example is given in Figure 8.6; a user has selected a tumour vertex within the querys graph visualisation to isolate similar tumour ROIs in the query and retrieved PET-CT images. A detailed description of the process is given by Algorithm 8.1.

The algorithm takes as inputs the selected vertex (v), the rank of the graph containing the vertex (graph), the vertex mappings (maps) generated by the graph comparison algorithm, and the database of ROIs (segments). The set of maps contains the best query vertex to indexed graph vertex mapping (φ) as calculated by the our graph comparison technique (Algorithm 6.2). This means that if graph is not the query then there will be no direct mapping from v to vertices of other non-query graphs. For this reason, the algorithm finds v_Q , the query vertex that mapped to v, and calls itself with the query vertex and graph as new inputs (lines 4 to 13). The nested call is able to use maps directly to find the mapped vertices; the function GETROI(v_S , segments) then retrieves the 3D boundary of the ROI represented by the mapped vertex v_S in segments, the database of ROIs (see Figure 5.1). The collection of ROIs (one for the query image, and one for every other retrieved image) is then returned.

Two special cases occur on lines 11 and 25. Line 11 deals with the case where the non-query vertex v is inserted into the query graph. In this case, there is no corresponding mapped vertex in the query graph ($v_Q = \emptyset$) nor any similar vertices in other retrieved graphs. As such, the only ROI highlighted is in the image corresponding to the selected graph. Line 25 deals with the case where

111	Southand Star Mapped Vertex Rol Doundary Extraction
1:	function EXTRACTBOUNDS $(v, graph, maps, segments)$
2:	$bounds \leftarrow \emptyset$
3:	$\mathbf{if} \ graph \neq 0 \ \mathbf{then}$
4:	$\varphi \leftarrow maps (graph)$
5:	for all $pair \in \varphi$ do
6:	if second element of $pair$ is v then
7:	$v_Q \leftarrow \text{first element of } pair$
8:	$\mathbf{if}v_Q\neq \varnothing\mathbf{then}$
9:	return EXTRACTBOUNDS $(v_Q, 0, maps, segments)$
10:	end if
11:	$bounds \leftarrow bounds \cup (graph, \text{GETROI}(v, segments))$
12:	end if
13:	end for
14:	else
15:	$bounds \leftarrow bounds \cup (0, \text{GETROI}(v, segments))$
16:	for $i = 1 \rightarrow$ length of maps do
17:	$\varphi \leftarrow maps\left(i\right)$
18:	for all $pair \in \varphi$ do
19:	if first element of $pair$ is v then
20:	$v_S \leftarrow$ second element of pair
21:	$bound \leftarrow \emptyset$
22:	$\mathbf{if}v_S\neq \varnothing\mathbf{then}$
23:	$bound \leftarrow \text{GETROI}(v_S, segments)$
24:	end if
25:	$bounds \leftarrow bounds \cup (i, bound)$
26:	end if
27:	end for
28:	end for
29:	end if
30:	return bounds
$31 \cdot$	end function

Algorithm 8.1 Mapped Vertex ROI Boundary Extraction

the query vertex v is inserted into a specific retrieved graph. In this case, there is no corresponding mapped query for that specific retrieved graph ($v_S = \emptyset$); there could be valid mappings for other retrieved graphs. As such, no ROI is highlighted in the image where the query vertex was inserted.

This interaction was only possible due to our storage of the segmented ROIs (see Figure 5.1) and the way the underlying graph comparison algorithm operated. The similarity of two images was calculated on the basis of the graph edit distance between their corresponding graph representations. The set of vertex edit operations can be reinterpreted as mappings between query vertices and their associated vertices in the retrieved images, i.e., it indicates for every query vertex (and its associated ROI) the most similar vertex (and associated ROI) of a retrieved graph (image) as determined by the similarity measurement algorithm. The interaction is a visual way of presenting these mappings.

Chapter 9

Evaluation of Retrieval User Interface

In this chapter, we present the evaluation of our method for retrieval visualisation and interpretation. We assessed our work through user evaluation of the interpretation capabilities of our retrieval UI, and measuring the computational performance of our system.

9.1 Evaluation Procedure

9.1.1 Materials

For our experiments we used the clinical PET-CT data set described in Section 7.1.3. Our preprocessing stage [196] cropped and scaled the axial images into the same coordinate space with a new resolution of 256 x 256 pixels at 1.95mm x 1.95mm. We also obtained the clinical reports for each of the studies.

Our system was implemented on an Intel Core 2 Quad CPU clocked at 2.40GHz with 4GB of RAM running Windows 7 64-bit. The underlying retrieval engine was implemented in MATLAB 2012a; the interface was implemented in

Java using ImageJ [200] and the Java Universal Network/Graph (JUNG) framework [201] as libraries.

9.1.2 Survey Procedure

We examined the capabilities of our UI through a set of user surveys, which were approved by our institutional human ethics committee. We recruited 10 students (age: mean 25.6, standard deviation 4.4, gender: 3 female, 7 male) from within our department (years studied: mean 5.6, standard deviation 2.8, 4 undergraduate, 6 graduate), who were otherwise uninvolved in this research project. The number of participants was sufficient to gain qualitative insights about the UI's capabilities [202]. The students were involved in medical image processing (e.g., segmentation, classification, registration, visualisation) or imagerelated research (e.g., telemedicine). Our purpose was to evaluate the various different capabilities of our UI instead of its application to a specific clinical task. Furthermore, for an initial evaluation and to elict comments for improvement we believed that the participants needed to have knowledge about the complexity of medical images, concepts of image search, and general multimedia processing. The students had the unique combination of skills that would allow an evaluation of all aspects of the system in relation to interpreting the retrieved images.

The participants were asked to complete a retrieval task using our proposed retrieval UI as well as a baseline UI, which had the same layout but only contained elements of typical CBIR UIs. That is, the baseline UI gave access to different image views and the clinical reports but there were no graph visualisations or graph interactions. Prior to undertaking the task, each participant was given a training session with a walkthrough of the functions of each UI. During the training session, users were also exposed to the types of information used by the retrieval system to calculate retrieved image rankings as well as the tools provided by the UI to explore and interpret the retrieved images. However, we assumed that in practice each user would have a different interpretation of what they consider to be similar. As such, image similarity was left to the subjective choice of individual participants using any criteria they wished (e.g., tumour location, the size and shape of structures, elements in the clinical report, glucose uptake, number of tumours, etc.). This allowed us to measure whether the UI let them achieve their own retrieval goals.

During the task, the participants performed an initial query using a PET-CT image with two lung tumours. The first tumour occurred within the left lung while the second tumour invaded the mediastinum from the left lung; the tumours had different volumes (this was not readily apparent from the image). In addition, the characteristics of the image (involvement of multiple organs and tumours, elements of different volumes) was typical of PET-CT images, meaning that the retrieved results would also have varied characteristics and would require the participants to study them from different perspectives (or use different UI tools) before they could decide whether an image was similar to the query. All participants used the same image as the initial query. After completing the task, the participants were given the opportunity to try different queries of their choice, some simpler, e.g., a single tumour, and some more complex, e.g., multiple tumours in multiple organs. The participants did not have to segment the ROIs in the query images as this had been performed in advance.

The participants used the same initial query for both the baseline and proposed system evaluation. This was done to ensure that the results of a different query did not influence their opinion of the different systems. We carried out tests in both sequences (baseline first then proposed, and vice versa) to demonstrate that human learning did not affect the responses received by the second system. Half of the participants were assigned to each sequence. Each user was given 7 minutes per task. The time limit was determined empirically according to the time needed to examine the cases retrieved by a query. The time limit also

#	Statement
1	The viewer was easy to use.
2	The viewer was fast and was responsive to input.
3	The viewer rendered good quality images.
4	The viewer provided all the controls I needed.
5	The viewer and its controls were laid-out well.
6	I was able to use the viewer to understand the images.
7	The viewer contained all information necessary to understand the images.
8	The amount of image and other information presented did not confuse me.
9	When searching, the viewer generally found the most similar images first.
10	I was able to use the search and filtering framework to improve my understanding of the image data.

Table 9.1: User Survey Statements

ensured that both systems (baseline and proposed) were used for the exact same period, allowing us to compare the effectiveness of each independent of the time spent on the task.

The participants completed an anonymous survey where they were asked about their experiences using the system. In the survey, the participants indicated their level of agreement or disagreement with the several statements using a 5-point scale (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree); this is a well-established approach for gathering information about user experiences [203]. We evaluated our approach by comparing the results of the survey for both the proposed and baseline systems. The statements in the survey are given in Table 9.1. The survey statements were designed to measure the UI and interaction requirements presented in Section 8.2. Participants were also allowed to leave free text comments on any aspect of the system they wished to mention.

9.1.3 Other Evaluation

We also examined the performance of our implementation by calculating the mean time to perform several common tasks. Specifically, we measured the time taken to perform a query, to browse to a new page of visualisations, and to swap to a different view. We also simulated increases in the database size to examine the scalability of the retrieval time compared to the database size

9.2 Results

9.2.1 Survey Responses

Figures 9.1 to 9.3 show the average responses to our survey across all participants. Figure 9.1 shows the combined responses across all participants. Figure 9.2 shows the responses of participants who used the baseline viewer first and then used our proposed viewer. Figure 9.3 shows the responses of participants who first used our proposed viewer and then used the baseline viewer. Table 9.2 summarises the median and range of the responses across all participants. The table also shows the significance (*p*-value) of the responses across the entire participant population calculated using the Wilcoxon rank sum. The circles in Figure 9.1 indicate statements where the differences in responses between the baseline and proposed system were statistically significant (p < 0.05).

9.2.2 Participant Comments

All participants provided textual feedback to explain their opinions and suggest improvements. The main positive comments about our proposed UI were summarised as follows:

• The proposed UI was much better than the baseline UI (4 comments).

- The abstractions were more convenient for understanding the images (2 comments).
- It was more difficult to locate small tumours or to see which images were similar using the baseline UI (2 comments).
- Without the graph abstraction in the baseline system, there was a high reliance on the fused images to determine image similarity (1 comments).

The main negative comments were summarised as follows:

- In the proposed system the graph edges in the abstractions could cause confusion if there was a lot of overlap, especially in a complex image (1 comments).
- The speed of operation for both systems was a little slow, particularly the time to load a new page of results (5 comments).

The participants also suggested several improvements. These included:

- The ability to hide or show different features of the UI.
- Displaying the clinical report within the panel of its associated image.
- The ability to magnify the images, thereby allowing users to view small ROIs in greater detail.
- A more intuitive method to edit the abstractions and the query.

9.2.3 Performance Evaluation

Table 9.3 lists the resource usage of the baseline and proposed retrieval UIs. Table 9.4 shows the time taken by standard tasks using the different UIs. We used a two-sample t-test to determine if the differences in resource usage and task time were statistically significance. Figure 9.4 shows the scalability of the system by comparing retrieval time to database size.

Statement	Baseline		Proposed		p-value
	Median	Range	Median	Range	
1	4	2	4.5	1	0.0334
2	3.5	3	3	3	0.6937
3	4	2	4	2	1.0000
4	3	2	4	1	0.0039
5	4	3	4	2	0.1104
6	4	2	4.5	2	0.0426
7	3	3	4	2	0.0128
8	3	2	4	2	0.0598
9	4	2	4	2	0.4253
10	4	2	4	2	0.0246

 Table 9.2: Survey Response Summary



Figure 9.1: Survey responses – all participants.

9.3 Discussion

The survey results (Figs. 6 to 8, Table 2) show that our proposed method achieved better responses in 8 out of 10 questions than the baseline UI. It is important

145



Figure 9.2: Survey responses – evaluating the baseline system first.



Figure 9.3: Survey responses – evaluating the proposed system first.

to note that for these statements, the proposed method equaled or outperformed the baseline method regardless of which system the users were exposed to first, as

	CPU (%)		Heap RA	AM (MB)	Non-Heap (MB)	
	Baseline	Proposed	Baseline	Proposed	Baseline	Proposed
Mean	6.0%	5.8%	222	246	22	23
Std.Dev.	5.9%	6.7%	125	122	2	3
Peak	31.1%	29.9%	472	499	23	26
<i>p</i> -value	0.9112		0.2077		0.0010)	

Table 9.3: Resource Usage

Table 9.4: Task Timing

	Times (ms)					
Tool	Baseline		Proposed			
Lask	Mean	Std.Dev.	Mean	Std.Dev.	<i>p</i> -value	
Switch view	59	77	92	168	0.4298	
Load new page	5032	2082	5161	2422	0.8575	
Query database (size: 50)	12299	2646	12877	2954	0.7528	



Figure 9.4: Retrieval time in relation to database size.

seen by the results in Figures 9.2 and 9.3. That is to say, the order in which the users evaluated the two systems did not seem to bias the final average results. The

survey responses and comments indicated that the capabilities of our proposed UI (multiple views, displaying abstractions and supplementary data, interactions for indicating similarity) improved the abilities of users to interpret the retrieved images.

According to the survey, both methods rendered images equally well (see the responses to Statement 3). This is understandable because there was no difference between the UIs in the manner in which the images were displayed.

The responses to Statement 1 indicate that the proposed UI was significantly easier to use than the baseline system, despite displaying more UI elements and providing more interactions to the user. The reason behind this is seen in the responses to Statement 4. The participants felt that the baseline viewing method did not provide them with the controls necessary to manipulate and completely understand the retrieved images; the difference in the responses was also statistically significant (p = 0.0039). As shown in Figure 9.1, the users felt that these controls were laid-out in a manner that allowed them to effectively view the data and interact with the system (Statement 5). A similar situation to Statement 4 occurs for the amount of information provided by the systems, and whether it was sufficient for understanding the images (Statement 7); this is discussed further in a later part of this section.

The baseline method only achieved a better survey response in regards to speed of operation (Statement 2). However, this result is not statistically significant (p = 0.6937). Upon further analysis, it can be shown that the difference in survey responses was entirely due to the group of participants who used the baseline system first. The users who were exposed to the proposed method first did not see any significant difference in the speed of operation between the two UIs; the mean and deviation in the response was the same for Statement 2 (as seen in Figure 9.3). Compared to the time taken to execute a query or to change a page, which are in the order of seconds, the graph display times are negligible (measured to be about 2ms). As such, we believe the low responses for Statement 2 for the proposed UI were probably due to apparent lag between the graph and image display when loading a new page. In the proposed system, all graphs were simultaneously displayed first and were then followed by the images, again simultaneously. The fast graph display made it seem like there was a delay in displaying the images (about 5 seconds), even though this same delay existed when loading a new page in the baseline system.

Statements 6 to 8 were used to measure whether users believed the user interface assisted them in interpreting the retrieved images. The proposed method achieved higher mean responses than the baseline method in all of these questions, regardless of which system the participants first used. In particular the responses to Statement 7 demonstrated that participants believed that the baseline method did not provide enough information for understanding the retrieved images when compared to the proposed method. The responses to Statement 8 demonstrated that the extra information provided by the graph visualisations and its associated interactions did not confuse the users.

Another finding was that participants who used the baseline version first stated that the proposed method was more accurate in its retrieval; this can be seen by the responses to Statement 9 in Figure 9.2. These responses occurred even though the underlying retrieval algorithm for both cases was the same. This finding indicates that the graph visualisations allowed users to grasp the similarity between images more easily; this conclusion is further supported by the significantly higher responses to Statement 10 (p = 0.0246), which indicated that the retrieved images assisted users in understanding the query image. Several participants commented that the graph abstractions made it easier for them to understand the images and the similarities between them. As such, we suggest that the capabilities provided by our UI can bridge the semantic gap by assisting users in understanding the similarities between complex images, such as PET-CT. The survey responses to Statements 3, 7, and 8 indicate that our proposed UI was capable of visualising multiple volumetric and multi-modality medical images, contained all information necessary for understanding the images (such as patient reports to indicate clinical aspects not apparent by visual inspection), and that the presentation of this information did not hinder the users. These outcomes are directly related to the visualisation recommendations stated earlier: we rendered multiple views that were of a good quality (Statement 3), we provided enough information to understand the images, by allowing users to browse through the data set and view patient reports for different images (Statement 7), and our use of the graph visualisation as an abstract aided in reducing confusion among users (Statement 8). Finally, our UI layout ensured that relevant information was presented in easy to understand (Statement 5) and easy to use (Statement 1) manner.

The outcomes for Statements 4, 6, and 10, indicate that our implementation of the recommended interactions provided the capabilities to allow users to understand the retrieved images. In particular, the controls on our UI allowed users to switch between different image views, to highlight similarities on the images, to rearrange the order of the presented data, and refine the query (Statement 4). These controls assisted the users in understanding the images (Statement 6). In particular, the responses to Statement 10 indicated that the combination of visualised data and the controls for interacting with and manipulating this data helped users in understanding the characteristics that made images similar. The comments from the participants also indicated that it was easier to understand the images using the proposed UI.

The capabilities of the proposed UI were directly responsible for the better responses received in the survey and the positive participant comments. For example, if the user's subjective search intent was based on the number of tumours, then this was directly evident in the graph abstraction; the user did not have to navigate through the image stacks to count the number of tumours. Similarly, image similarity based on the anatomical location of tumours could also be determined from the graph abstractions; the filter options provided another way to narrow search based on tumour location. As another example, when searching for a primary tumour of a similar volume, the user could select the vertex corresponding to the tumour, obtain the volume feature, and then use the search tools to set upper and lower bounds (e.g., $\pm 10\%$ of the measured volume) on the volume of tumours. These functionalities are among the reasons why the proposed UI was more convenient to use.

The interactions that visually indicated image similarity also helped to save time during interpretation. The user did not have to manually navigate through each individual view of each image (3 coronal views per image) to find the region of their choice. A single click automatically aligned all coronal stacks of all the retrieved images to middle slice of the appropriate ROI. The outline of the ROI was marked on each slice to indicate its limits (so the user did not need to navigate beyond it).

We also suggest that the capabilities of our UI can be optimised for clinical tasks. The UI has the potential to allow users to quickly isolate a subset of a large set of retrieved images. The abstractions can be used to eliminate cases that the user is not interested in without needing to examine the images in detail. We propose that physicians could adopt a similar UI for evidence-based diagnosis. After executing a query and applying any interactions they deem necessary, physicians could use the clinical reports of the retrieved cases to determine their diagnosis. The potential benefits of decision support based on retrieved information has already been discussed in the literature [22, 25].

The performance results demonstrate that the only notable difference between the proposed and baseline UIs was the use of non-heap RAM. However, the absolute difference in peak non-heap RAM usage (3 MB) is inconsequential given modern hardware. With a mean CPU usage of less than 6% and a mean heap RAM usage below 250 MB, our method has relatively low resource consumption. The peak CPU usage of less than 30% and a peak heap RAM usage of less than 500 MB indicates that our UI can be quite effectively used on modern consumer hardware.

Table 9.4 shows that most common tasks were performed in real time. Switching between views (PET, CT, fused, MIP) was virtually instantaneous with a mean time of 92ms. Loading a full page of new results (4 new PET-CT images, each with multiple views) took on average 5.16s. Querying the database was the most time-consuming task. The mean retrieval time for our implementation was 12.88s on a database of 50 PET-CT images. Our evaluation of the system's scalability showed that the retrieval time scaled linearly with database size (as indicated by the line of best fit in Figure 9.4), suggesting that optimisations would be helpful for practical use with large image repositories. Most users were neutral about the speed and responsiveness of the proposed method (Statement 2), suggesting that it was acceptable for their purpose but that speed improvements would be helpful. Our retrieval algorithm (described in Chapter 6) does not require a specific implementation; as such, in the future we can optimise the query time through the use of parallel computing or cloud-based hardware.

Chapter 10

Discussion

This chapter discusses the novel capabilities introduced in this thesis and the future work that will overcome their limitations.

10.1 Capabilities

The CBIR framework presented in this thesis was designed to enable the retrieval of multi-modality medical images like PET-CT by exploiting the complementary information provided by each image modality. The design of our features, graph representation, and similarity measurement algorithm was inspired by clinical cancer classification schemes [171, 172], and their use of geometric and topologic information (such as the relationships between tumours and anatomy) for disease staging. Our methods have the following capabilities.

10.1.1 Structural Extensibility

Graphs are not constrained to a fixed order or size [179]. As such, the order of the CAPP graph can change to match the number of ROIs in the image while the size of the graph can be altered according to the relationships in the pattern. Our pruning function (Equation 6.1) can be implemented or adapted to define edges between any extra vertices that have been created.

As such, one of the major capabilities of our framework is the ability to represent images both on a regional level (such as thorax in lung cancer imaging) to whole-body level (for diseases like lymphoma). Improvements in image acquisition and segmentation technologies (such as whole-body segmentation algorithms [204]) will enable the extraction of even finer details and as well as the delineation of a larger number of ROIs. Our graphs will be able to represent this new information.

10.1.2 Modality Extensibility

In this thesis, we have only dealt with dual-modality images and have evaluated our work using clinical PET-CT. However, our framework is extensible, only requiring modality-specific adjustments. Our approach can be optimised for multimodality images such as PET-MR or SPECT-CT through the use of appropriate registration, segmentation, and feature extraction algorithms.

Furthermore, our division of the entire vertex set V_K into anatomical (V_A) and pathology (V_P) subsets is not a limitation to the types of information that can be indexed on our graphs. The vertex set can be divided in other ways and could possibly contain more than two subsets. In this case, a graph with the vertex set $V_K = V_1 \cup V_2 \cup \cdots \cup V_n$ would have an associated feature alphabet $f = (f_1, f_2, \ldots, f_n).$

10.1.3 Feature Extensibility

In this thesis, we used feature sets that included standard geometric features [121] and complemented them with modality-specific information, e.g., CT texture and PET SUV for our PET-CT data set (Section 7.1.3). We also only indexed three types of features (Section 6.3.1).

However, graph representations can index any feature as an attribute of a vertex or an edge [121]. As such, it is possible to include other features among those that we have used already. The feature set should reflect the characteristics of the data set, and by not using a fixed feature set we are not restricting the use of our method with other images or for other applications.

Our feature normalisation and similarity measurement schemes have been designed for our application to PET-CT. Expanding the types of features (e.g., including free text labels) may require a modification to the normalisation algorithm (Appendix A) or the cost function (Equation 6.8).

10.1.4 Full Representation - Complete Graphs

Complete graphs represented every ROI and every relationship between ROIs. They essentially attempted to index all the information necessary for representing even the smallest detail in the images. However, such a representation came at a cost; there was no difference in structure between complete graphs with the same number of vertices, thereby reducing their ability to discriminate between cases where structure is important, e.g., images with tumours within different anatomical structures (as seen in Chapter 7).

Complete graphs may be useful when structure is not as important, such as when the tumours are already known to be within a particular structure, or in images where structure or arrangement is less important. They may also be useful in images where structures are generally the same but contain many minor individual feature variations across different ROIs.

10.1.5 Structural Representation - CAPP Graphs

Our CAPP graphs were designed to constrain tumours to spatially related anatomical structures. As such, they emphasised the spatial structure of the objects in the image, particularly the relative location of tumours in relation to anatomy. However, the representation of the anatomy in a CAPP graph (consisting of V_A and the edges between elements of V_A) forms a complete subgraph of the entire CAPP graph because most humans have the same arrangement of anatomical objects with differences in attributes. This was done in order to represent minor anatomical features that may be important for inter- and intra-patient similarity matching. This means that the CAPP graph was equivalent to complete graphs for healthy patients, e.g., images with no tumours. Its discriminatory power was higher when tumours were present within an image.

CAPP graphs were most useful when the images in the data set contained a wide variety of patients and disease patterns. This is ideal for clinical data where there are large naturally occurring variations among patients, and even images with similar diagnoses can have different structures and features.

10.2 Future Research Directions

Several new research directions are now possible due to the graph-based multimodality image retrieval methods introduced in this thesis. These research directions both leverage the strengths of our work but also build upon them to innovate in new areas. They also act to address the limitations of the thesis. While several research directions are seemingly obvious (application to PET-MR, introduction of new features, clinical evaluation, etc.), others form detailed areas of study with the potential for significant theoretical and clinical impact.

10.2.1 Probabilistic Graph Construction

The current CAPP graph construction technique is based upon the values of image features extracted from the ROIs. In particular, the pruning function is reliant upon the accurate segmentation of different structures. As such a more realistic approach would be a probabilistic pruning function that accounts for the accuracy of the segmentation algorithm and the features. In effect, the pruning function chooses the edge e that maximises the probability that it occurred with a given proximity value x, i.e., maximises P(e|proximity = x).

10.2.2 Including Temporal Information

In this thesis, we retrieved individual studies of patients, i.e., images acquired at a single time point. However, it is common for multiple scans of a single patient to be acquired during treatment. These images are acquired at different time points during the treatment cycle and allow physicians to monitor a particular patient's response to treatment. It allows physicians to intervene early if a particular treatment course (e.g., chemotherapy) is not performing as expected. As such, it is important to be able to retrieve similar *cases* for image-driven evidence-based diagnosis. The similarity of cases would be defined not just by disease location but also by the pattern of progression over time, i.e., stable disease, responding disease, or progressing disease. This would require the inclusion of temporal relationships, edges between the same vertex representing how it has evolved with time.

10.2.3 Improved Segmentation and Registration

In our experiments, we have shown that current segmentation and hardware registration algorithms were sufficient for the PET-CT lung tumour data set. It is important to note that the performance and capabilities of our retrieval algorithm will improve alongside advances in segmentation and registration algorithms. In addition, improvements to imaging technology that benefit segmentation and registration (such as acquiring higher resolution PET images) will also inherently improve the capabilities of our retrieval framework. This is because our graph construction technique relies upon segmentation to enable the extraction of ROI features. Similarly, registration enables the extraction of the spatial relationships between ROIs in different modalities.

10.2.4 Optimal Feature Selection

In this thesis, we adapted the geometric and spatial features described in [121] and complemented them with modality specific information, such as SUV for PET or texture for CT; the reasons for this were stated in Section 6.3.3. The retrieval performance could be improved by learning an optimal set of features. However, standard approaches to feature selection [205] cannot be directly applied to our graph representations.

A new method for feature selection is needed, one which can balance structural information as well as the image features indexed on vertices and edges. The optimal set of features will therefore be subgraphs of the current CAPP graphs. Each of these subgraphs will have their own individual feature sets (vertex and edge attributes) that optimise the relative importance of the structure modelled by that subgraph.

10.2.5 Reduction in Computational Complexity

One of the key limitations of graph edit distance based similarity matching is the high computational complexity of graph algorithms [179]. Increasing the order of the graph representations will eventually cause the graph comparison algorithm to no longer return results in an appropriate time frame. This introduces a trade-off between accuracy and time [121]. In our case this trade-off could be implemented by reducing the beam size to reduce the search space (see Section 7.3.4). However, this is only a delaying strategy since gains from adjusting the beam size are linear (see Figure 7.11) while the time increases exponentially as the graph order

increases (see Figure 7.12).

Alternate approaches to graph similarity matching are therefore necessary. Potential solutions include approaches based on kernel machines [176, 179, 192, 194] or polynomial time algorithms for graph edit distance approximations [206]. In these cases, the algorithms and results in this thesis provide a baseline for evaluating the viability of future approaches.

10.2.6 Designing an A^* Heuristic Function

The A^{*} algorithm uses a heuristic function h(n) to determine the order in which the search visits the nodes in the search tree. The function h(n) is an estimate of the distance from the search tree node n to the goal state; an admissible h(x)never overestimates this distance [186]. The time complexity of A^{*} is dependent upon this heuristic, with the worst-case being h(n) = 0. The inclusion of multiple modalities and feature sets on our graphs offers the opportunity to derive a h(n)that is optimised for our particular retrieval task.

10.2.7 Vector Space Embedding

Our graph comparison algorithm (Section 6.6) is based upon the graph-edit distance. As such, this offers the opportunity to investigate vector space embeddings [207] of our graphs. A vector space embedding of our graphs would allow for the application of a range of methods and tools available for vectors, e.g., support vector machines and feature selection algorithms. One major challenge will be deriving a method to select the prototype graphs in a manner that accounts for the variation in anatomical features, tumour features, and the graph structure. Vector space embeddings could also potentially be used to optimise retrieval time. We have already begun research efforts in this area [140]. However, an inherent disadvantage of embeddings is that a direct vertex-tovertex mappings will not be available for every graph in the data set. As such, a new method for visually indicating similar regions will be required for our UI.

10.2.8 Hardware-Driven Graph Matching

The future work described Sections 10.2.5, 10.2.6, and 10.2.7 all serve to improve the computational performance through the use of more efficient algorithms, i.e., they are all software based. Hardware-based solutions could also improve the computational performance of our algorithm. In particular, field programmable gate arrays (FPGAs) could be used to design a hardware of our graph matching algorithm [208]. Parallelisation will also improve the throughput of our CBIR method.

10.2.9 Data Mining and Classification

Our structural representation of relationships between tumours and organs offers the opportunity for other image similarity research, such as data mining. In particular, we believe the use of frequent subgraph mining (FSM) [209] on a large collection of CAPP graphs can reveal information about tumour occurrences, such as whether commonly occurring tumour localisation patterns have reoccurring ROI characteristics. We propose to expand FSM algorithms to account for graph attributes extracted from different modalities.

10.2.10 Improved Visualisation Techniques

The UI presented in Chapter 8 utilises preprocessed views of PET-CT data. As such, the user's control over the views is limited entirely to the preprocessed images. While the use of preprocessed images does offer benefits for real-time execution it also places limitations on the ways in which the retrieval UI can
be used. Furthermore, there is an underlying assumption that the preprocessed images are sufficient to interpret the entire image.

The integration of advanced 3D rendering techniques, such as direct volume rendering [210], should be investigated as a potential replacement for 2D slice based views of tomographic data. Furthermore, the user should be provided with more advanced control over the visualisations, such as the ability to adjust the image grey-level range (through the manipulation of window width and length), zoom functionality, and the ability to apply appropriate colour look up tables. However, care must be taken to ensure that these additional tools do not degrade the UIs real-time operational capabilities.

10.2.11 Integration of Visual Analytics

Our retrieval UI allows users to understand the similarity between images by visually indicating similar ROIs on the images. However, this does not describe the contribution of individual features. That is, while the retrieved images are ranked according to their overall similarity, there is no method by which users can examine how individual features affected the overall similarity. The field of visual analytics [211] provides techniques for analytical reasoning and visual representation of analytical results; we believe that integrating such technologies into a CBIR UI has the potential to provide even greater insight to users attempting to interpret retrieved images.

Chapter 11

Conclusions

In this thesis, we have presented a graph-based method for the retrieval of multi-modality medical images. Our formulation of the graph representation of medical images emphasised the spatial relationships between tumours and organs; this representation in turn emphasised disease localisation during retrieval. Our method introduced image similarity based upon a combination of modalityspecific features and structure between ROIs extracted from two different modalities.

Furthermore, we also presented the design of a UI to enable effective interpretation of retrieved PET-CT images. In addition to visualising multiple 3D volumes, our UI provided supplementary information to assist user interpretation. This supplementary information was in the form of clinical reports associated with the images as well as graph abstractions that acted as a summary of the complex image information. We also defined a new interaction that visually indicated similar regions in images thereby allowing users to more easily locate and compare similar regions.

Our results demonstrated that our graph-based retrieval algorithm introduced the capability to search for multi-modality images on the basis of disease localisation. They results also showed that our CAPP graph, which emphasised the relationships between a tumour and its spatially nearest anatomical structures, had a higher retrieval precision than the complete graph, which represented all relationships. Our UI evaluation revealed that the graph visualisations and interactions in our proposed interface improved user's understanding of the retrieved images. In summary, this thesis has advanced the state-of-the-art by introducing a novel graph-based approach for the retrieval of multi-modality medical images. We suggest that our approach can be adapted to other applications where knowledge must be retrieved and visualised from repositories containing multi-modality information. Appendices

Appendix A

Graph Feature Normalisation

This appendix provides detailed algorithms for conducting our feature normalisation process. A summary of the process was provided in Section 6.4.

A.1 Index Normalisation

Algorithms A.1 describes the process for normalising the graph index (collection of graphs). The first stage (line 3) of index normalisation involves finding the distribution (mean and standard deviation) of each feature across the entire index. The distribution of features is calculated separately for each modality. This process is described in Section A.2. The distribution of features is then used to normalise each graph (line 5), as described in Section A.3.

Algorithm A.1 Graph Index Normalisation			
1: function NORMALISEINDEX(<i>index</i> , f_A , f_P , f_S)			
2: $normIndex \leftarrow \emptyset$			
3: $(F_{\mu}, F_{\sigma}) \leftarrow \text{ACCUMULATE}(index, f_A, f_P, f_S)$			
4: for all $G_S \in index$ do			
5: $normGraph \leftarrow NORMALISEGRAPH(G_S, F_{\mu}, F_{\sigma})$			
6: $normIndex \leftarrow normIndex \cup normGraph$			
7: end for			
8: return $(normIndex, F_{\mu}, F_{\sigma})$			
9: end function			

A.2 Calculating Feature Distributions

To calculate the feature distribution, we first accumulate all measurement features on the graph vertices and edges. This is performed for each modality separately. The process is described by Algorithms A.2 to A.4.

The accumulation creates three sets F_A , F_P , and F_S , each of which contain pairs F = (f, X) where f is the feature name and X is a set that contains *all* occurrences of that feature among all graph elements (vertex or edge) of the same modality in the entire data set. These pairs are used by Algorithm A.5 to calculate the mean and standard deviation of these features, again separately for each modality.

Alg	Algorithm A.2 Graph Feature Accumulation				
1:	1: function ACCUMULATE(<i>index</i> , f_A , f_P , f_S)				
2:	$F_A \leftarrow \text{INITIALISE}(f_A)$				
3:	$F_P \leftarrow \text{INITIALISE}(f_A)$				
4:	$F_S \leftarrow \text{INITIALISE}(f_A)$				
5:	for all $G_S \in index$ do				
6:	$V \leftarrow V\left(G_S\right)$				
7:	$E \leftarrow E\left(G_S\right)$				
8:	for all $v \in V$ do \triangleright Accumulate vertex features				
9:	if $MODALITY(v) = A$ then				
10:	$F_A \leftarrow \text{ACCUMULATEFORONE}(f_A, F_A, v) \mathrel{\triangleright} \text{Anatomical Vertex}$				
11:	else				
12:	$F_P \leftarrow \text{ACCUMULATEFORONE}(f_P, F_P, v) \qquad \triangleright \text{ Tumour Vertex}$				
13:	end if				
14:	end for				
15:	for all $e \in E$ do \triangleright Accumulate edge features				
16:	$F_S \leftarrow \text{ACCUMULATEFORONE}(f_S, F_S, e)$				
17:	end for				
18:	end for				
19:	$(F_{A\mu}, F_{A\sigma}) \leftarrow \text{DISTRIBUTIONOF}(f_A, F_A)$				
20:	$(F_{P\mu}, F_{P\sigma}) \leftarrow \text{DISTRIBUTIONOF}(f_P, F_P)$				
21:	$(F_{S\mu}, F_{S\sigma}) \leftarrow \text{DISTRIBUTIONOF}(f_S, F_S)$				
22:	$F_{\mu} \leftarrow (F_{A\mu}, F_{P\mu}, F_{S\mu})$				
23:	$F_{\sigma} \leftarrow (F_{A\sigma}, F_{P\sigma}, F_{S\sigma})$				
24:	$\mathbf{return} \ (F_{\mu}, F_{\sigma})$				
25:	e end function				

```
Algorithm A.3 Feature Collection Initialisation
```

1: function INITIALISE(f_{α}) $F \leftarrow \emptyset$ 2: for all $f \in f_{\alpha}$ do 3: if ISNUMERICFEATURE(f) then 4: $F \leftarrow F \cup (f, \emptyset)$ 5: end if 6: 7:end for return F8: 9: end function

Algorithm A.4 Graph Element Feature Accumulation

1: function ACCUMULATEFORONE $(f_{\alpha}, F_{coll}, elem)$ 2: for all $f \in f_{\alpha}$ do if ISNUMERICFEATURE(f) then 3: $F \leftarrow f(F_{coll})$ $\triangleright F = (f, X)$ where X is a set of feature values 4: $F_{coll} \leftarrow F_{coll} \setminus F$ \triangleright Remove feature entry from collection 5: $X \leftarrow X(F)$ 6: $x \leftarrow f(elem)$ 7: \triangleright Obtain feature value $X \leftarrow X \cup x$ \triangleright Expand set of feature values 8: $F \leftarrow (f, X)$ 9: $F_{coll} \leftarrow F_{coll} \cup F$ \triangleright Reinsert expanded entry 10: end if 11: end for 12:return F_{coll} \triangleright Return updated collection 13:14: end function

Algorithm A.5 Feature Distribution Calculation

1: function DISTRIBUTIONOF (f_{α}, F_{coll}) 2: $F_{\mu} \leftarrow \emptyset$ $F_{\sigma} \leftarrow \emptyset$ 3: 4: for all $f \in f_{\alpha}$ do if ISNUMERICFEATURE(f) then 5: $F \leftarrow f(F_{coll})$ 6: $X \leftarrow X(F)$ 7: $\mu_f \leftarrow \text{MEAN}(X)$ 8: $\sigma_f \leftarrow \operatorname{STD}(X)$ 9: 10: $F_{\mu} \leftarrow F_{\mu} \cup (f, \mu_f)$ $F_{\sigma} \leftarrow F_{\sigma} \cup (f, \sigma_f)$ 11: end if 12:end for 13:return (F_{μ}, F_{σ}) 14: 15: end function

A.3 Graph Normalisation

Our graph normalisation process is described by Algorithms A.6 and A.7. The normalisation process is dependent upon not only the modality from which a feature was extracted but also the type of the feature (as described in Section 6.3.1).

Alg	Algorithm A.6 Graph Normalisation			
1:	function NORMALISEGRAPH (G, F_{μ}, F_{σ})			
2:	\triangleright Note: $F_{\mu} = (F_{A\mu}, F_{P\mu}, F_{S\mu})$ and $F_{\sigma} = (F_{A\sigma}, F_{P\sigma}, F_{S\sigma})$			
3:	$^{N}V \leftarrow \emptyset$			
4:	$^{N}E \leftarrow \emptyset$			
5:	$V \leftarrow V(G)$			
6:	$E \leftarrow E(G)$			
7:	$(f_A, f_P, f_S) \leftarrow f(G)$			
8:	for all $v \in V$ do			
9:	if $MODALITY(v) = A$ then			
10:	$^{N}v \leftarrow \text{NORMALISEELEMENT}(v, f_{A}, F_{A\mu}F_{A\sigma})$			
11:	$^{N}V \leftarrow^{N} V \cup^{N} v$			
12:	else			
13:	$^{N}v \leftarrow \text{NORMALISEELEMENT}(v, f_{P}, F_{P\mu}F_{P\sigma})$			
14:	$^{N}V \leftarrow^{N} V \cup^{N} v$			
15:	end if			
16:	end for			
17:	for all $e \in E$ do			
18:	$^{N}e \leftarrow \text{NORMALISEELEMENT}(e, f_{S}, F_{S\mu}F_{S\sigma})$			
19:	${}^{N}E \leftarrow {}^{N}E \cup {}^{N}e$			
20:	end for			
21:	end function			

Measurement features were normalised by the function *normalise* (called on line 8 of Algorithm A.7 and given by Equation A.1); this function was directly adapted from [87]. It linearly scaled a measurement value x to a random variable with zero mean and unit variance, ensuring 99% of all values were normalised to the range the range [0, 1]. As in [87], values outside this range were shifted to the closest value (0 or 1). The function was defined as follows:

normalise
$$(x, f_{\mu}, f_{\sigma}) = \frac{(x - \mu_f)/3\sigma_f + 1}{2}$$
 (A.1)

Algorithm A.7 Vertex and Edge Normalisation				
1:	1: function NORMALISEELEMENT($elem, f_{\alpha}, F_{\mu}, F_{\sigma}$)			
2:	2: $^{N}elem \leftarrow \emptyset$			
3:	3: for all $f \in f_{\alpha}$ do			
4:	$4: \qquad x \leftarrow f(elem)$			
5:	5: if ISNUMERICFEATURE (f) then			
6:	$f_{\mu} \leftarrow f(F_{\mu})$			
7:	$f_{\sigma} \leftarrow f(F_{\sigma})$			
8:	8: $\tilde{x} \leftarrow normalise(x, f_{\mu}, f_{\sigma})$			
9:	9: else if $ISANGLEFEATURE(f)$ then			
10:	$S: \qquad sVal \leftarrow \frac{\sin x + 1}{4}$			
11:	1: $cVal \leftarrow \frac{\cos x + 1}{4}$			
12:	$2: \qquad \qquad ^{N}elem \leftarrow {}^{N}elem \cup sVal \cup cVal$			
13:	3: else			
14:	4: $^{N}elem \leftarrow ^{N}elem \cup x \qquad \rhd \text{Point}$	t sets are unchanged		
15:	5: end if			
16:	6: end for			
17:	7: return ^{N} $elem$			
18:	8: end function			

where x was the actual feature value, and μ_f and σ_f were the mean and standard deviation of the feature in the data set.

Angular values were normalised by function of their sine and cosine values, as in [121]. This can be seen on lines 10 and 11 of Algorithm A.7. This normalisation method avoided discrepancies near the extreme angles $(\pm (\pi - \epsilon))$, for a small value $\epsilon > 0$).

We did not normalise point set features, instead electing to normalise the distance between two point set features during similarity measurement. That is, we normalised the difference between two point sets such that the distance value ranged from 0 (total similarity) to 1 (total dissimilarity). This was achieved using the Jaccard distance:

$$d_J(q_{pts}, s_{pts}) = 1 - \frac{|q_{pts} \cap s_{pts}|}{|q_{pts} \cup s_{pts}|}$$
(A.2)

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

References

- K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4–5, pp. 198–211, 2007.
- [2] H. Zaidi, H. Vees, and M. Wissmeyer, "Molecular PET/CT imaging-guided radiation therapy treatment planning," *Academic Radiology*, vol. 16, no. 9, pp. 1108–1133, 2009.
- [3] C. D. Marcus, V. Ladam-Marcus, C. Cucu, O. Bouché, L. Lucas, and C. Hoeffel, "Imaging techniques to evaluate the response to treatment in oncology: Current standards and perspectives," *Critical Reviews in Oncology/Hematology*, vol. 72, no. 3, pp. 217–238, 2009.
- [4] D. W. Townsend and T. Beyer, "A combined PET/CT scanner: the path to true image fusion," *British Journal of Radiology*, vol. 75 (Supplement 9), pp. S24–S30, 2002.
- [5] D. W. Townsend, T. Beyer, and T. M. Blodgett, "PET/CT scanners: A hardware approach to image fusion," *Seminars in Nuclear Medicine*, vol. 33, no. 3, pp. 193–204, 2003.
- [6] M. S. Judenhofer, C. Catana, B. K. Swann, S. B. Siegel, W.-I. Jung, R. E. Nutt, S. R. Cherry, C. D. Claussen, and B. J. Pichler, "PET/MR images acquired with a compact MR-compatible PET detector in a 7-T magnet," *Radiology*, vol. 244, no. 3, pp. 807–814, 2007.
- [7] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications—clinical benefits and future directions," *International Journal of Medical Informatics*, vol. 73, no. 1, pp. 1–23, 2004.
- [8] H. Müller, X. Zhou, A. Depeursinge, M. Pitkanen, J. Iavindrasana, and A. Geissbuhler, "Medical visual information retrieval: State of the art and challenges ahead," in *IEEE International Conference on Multimedia and Expo*, pp. 683–686, 2007.
- [9] H. Müller, J. Kalpathy-Cramer, B. Caputo, T. Syeda-Mahmood, and F. Wang, "Overview of the first workshop on medical content-based retrieval for clinical decision support at MICCAI 2009," in *Medical Content-Based Retrieval for Clinical Decision Support* (B. Caputo, H. Müller,

T. Syeda-Mahmood, J. Duncan, F. Wang, and J. Kalpathy-Cramer, eds.), vol. 5853 of *Lecture Notes in Computer Science*, pp. 1–17, Springer Berlin Heidelberg, 2010.

- [10] H. K. Huang, PACS and Imaging Informatics: Basic Principles and Applications. Wiley & Sons, 2004.
- [11] W. D. Bidgood and S. C. Horii, "Introduction to the ACR-NEMA DICOM standard.," *Radiographics*, vol. 12, no. 2, pp. 345–355, 1992.
- [12] H. K. Huang, "Utilization of medical imaging informatics and biometrics technologies in healthcare delivery," *International Journal of Computer As*sisted Radiology and Surgery, vol. 3, pp. 27–39, 2008.
- [13] A. Rosset, O. Ratib, A. Geissbuhler, and J.-P. Vallée, "Integration of a multimedia teaching and reference database in a PACS environment," *Radiographics*, vol. 22, no. 6, pp. 1567–1577, 2002.
- [14] S. J. Doran, J. d'Arcy, D. J. Collins, R. Andriantsimiavona, M. Orton, D.-M. Koh, and M. O. Leach, "Informatics in radiology: Development of a research pacs for analysis of functional imaging data in clinical research and clinical trials," *Radiographics*, vol. 32, no. 7, pp. 2135–2150, 2012.
- [15] S. Sedghi, M. Sanderson, and P. Clough, "How do health care professionals select medical images they need?," *Aslib Proceedings*, vol. 64, no. 4, pp. 437– 456, 2012.
- [16] H. D. Tagare, C. C. Jaffe, and J. Duncan, "Medical image databases: A content-based retrieval approach," *Journal of the American Medical Informatics Association*, vol. 4, no. 3, pp. 184–198, 1997.
- [17] T. M. Lehmann, M. O. Guld, C. Thies, B. Fischer, D. Keysers, M. Kohnen, H. Schubert, and B. B. Wein, "Content-based image retrieval in medical applications for picture archiving and communication systems," in *Proceedings of SPIE 5033* (H. K. Huang and O. M. Ratib, eds.), pp. 109–117, 2003.
- [18] M. O. Gueld, M. Kohnen, D. Keysers, H. Schubert, B. B. Wein, J. Bredno, and T. M. Lehmann, "Quality of DICOM header information for image categorization," in *Proceedings of SPIE 4685* (E. L. Siegel and H. K. Huang, eds.), pp. 280–287, 2002.
- [19] D. Channin, P. Mongkolwat, V. Kleper, K. Sepukar, and D. Rubin, "The caBIGTM annotation and image markup project," *Journal of Digital Ima*ging, vol. 23, no. 2, pp. 217–225, 2010.
- [20] Y. Zalmanovitch and D. R. Vashdi, "Shrinking budgets, improving care: Trade-offs are unavoidable," BMJ, vol. 340, 2010.

- [21] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Contentbased image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000.
- [22] A. M. Aisen, L. S. Broderick, H. Winer-Muram, C. E. Brodley, A. C. Kak, C. Pavlopoulou, J. Dy, C.-R. Shyu, and A. Marchiori, "Automated storage and retrieval of thin-section CT images to assist diagnosis: System description and preliminary assessment," *Radiology*, vol. 228, no. 1, pp. 265–270, 2003.
- [23] H. Müller, A. Rosset, A. Garcia, J.-P. Vallée, and A. Geissbuhler, "Benefits of content-based visual data access in radiology," *Radiographics*, vol. 25, no. 3, pp. 849–858, 2005.
- [24] A. Rosset, H. Müller, M. Martins, N. Dfouni, and J.-P. Vallée, "Casimage project: A digital teaching files authoring environment," *Journal of Thoracic Imaging*, vol. 19, no. 2, pp. 103–108, 2004.
- [25] S. A. Napel, C. F. Beaulieu, C. Rodriguez, J. Cui, J. Xu, A. Gupta, D. Korenblum, H. Greenspan, Y. Ma, and D. L. Rubin, "Automated retrieval of CT images of liver lesions on the basis of image similarity: Method and preliminary results," *Radiology*, vol. 256, no. 1, pp. 243–252, 2010.
- [26] T. M. Blodgett, C. C. Meltzer, and D. W. Townsend, "PET/CT: Form and function," *Radiology*, vol. 242, no. 2, pp. 360–385, 2007.
- [27] T. Deserno, S. Antani, and R. Long, "Ontology of gaps in content-based image retrieval," *Journal of Digital Imaging*, vol. 22, no. 2, pp. 202–215, 2009.
- [28] H. K. Genant, C. Y. Wu, C. van Kuijk, and M. C. Nevitt, "Vertebral fracture assessment using a semiquantitative technique," *Journal of Bone* and Mineral Research, vol. 8, no. 9, pp. 1137–1148, 1993.
- [29] F. Hussain and K. L. P. Berrios, "Recommendation of panoramic radiography as a standard diagnostic tool for the head and neck region for the third molar cases: presentation of two distomolar cases," *European Journal* of Radiology Extra, vol. 74, no. 2, pp. e31–e35, 2010.
- [30] E. D. Pisano, C. Gatsonis, E. Hendrick, M. Yaffe, J. K. Baum, S. Acharyya, E. F. Conant, L. L. Fajardo, L. Bassett, C. D'Orsi, R. Jong, and M. Rebner, "Diagnostic performance of digital versus film mammography for breastcancer screening," *New England Journal of Medicine*, vol. 353, no. 17, pp. 1773–1783, 2005.
- [31] A. D. A. Maidment, "The future of medical imaging," Radiation Protection Dosimetry, vol. 139, no. 1-3, pp. 3–7, 2010.

- [32] G. T. Herman, Fundamentals of Computerized Tomography: Image Reconstruction from Projections. Advances in Computer Vision and Pattern Recognition, Springer-Verlag London, 2 ed., 2009.
- [33] Z. Abramson, S. Susarla, M. August, M. Troulis, and L. Kaban, "Threedimensional computed tomographic analysis of airway anatomy in patients with obstructive sleep apnea," *Journal of Oral and Maxillofacial Surgery*, vol. 68, no. 2, pp. 354–362, 2010.
- [34] A. Madani, C. Keyzer, and P. Gevenois, "Quantitative computed tomography assessment of lung structure and function in pulmonary emphysema," *European Respiratory Journal*, vol. 18, no. 4, pp. 720–730, 2001.
- [35] M. Hasegawa, S. Sone, S. Takashima, F. Li, Z. G. Yang, Y. Maruyama, and T. Watanabe, "Growth rate of small lung cancers detected on mass CT screening.," *British Journal of Radiology*, vol. 73, no. 876, pp. 1252–1259, 2000.
- [36] D. F. Yankelevitz, A. P. Reeves, W. J. Kostis, B. Zhao, and C. I. Henschke, "Small pulmonary nodules: Volumetrically determined growth rates based on CT evaluation," *Radiology*, vol. 217, no. 1, pp. 251–256, 2000.
- [37] P. E. Valk, D. Delbeke, D. L. Bailey, D. W. Townsend, and M. N. Maisey, eds., *Positron Emission Tomography: Clinical Practice*. Springer-Verlag London, 2006.
- [38] S. Belhassen and H. Zaidi, "A novel fuzzy c-means algorithm for unsupervised heterogeneous tumor quantification in pet," *Medical Physics*, vol. 37, no. 3, pp. 1309–1324, 2010.
- [39] J. Czernin, M. Dahlbom, O. Ratib, and C. Schiepers, Atlas of PET/CT Imaging in Oncology. Springer-Verlag Berlin, Heidelberg, 2004.
- [40] H. Zaidi and I. El Naqa, "PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, pp. 2165–2187, 2010.
- [41] J. A. Thie, "Understanding the standardized uptake value, its methods, and implications for usage," *Journal of Nuclear Medicine*, vol. 45, no. 9, pp. 1431–1434, 2004.
- [42] R. R. Edelman and S. Warach, "Magnetic resonance imaging," The New England Journal of Medicine, vol. 328, no. 10, pp. 708–716, 1993.
- [43] G. K. von Schulthess, H. C. Steinert, and T. F. Hany, "Integrated PET/CT: Current applications and future directions," *Radiology*, vol. 238, no. 2, pp. 405–422, 2006.

- [44] G. W. Goerres, G. K. von Schulthess, and H. C. Steinert, "Why most PET of lung and head-and-neck cancer will be PET/CT," *Journal of Nuclear Medicine*, vol. 45 (Supplement 1), pp. 66S–71S, 2004.
- [45] H. Zaidi, O. Mawlawi, and C. G. Orton, "Simultaneous PET/MR will replace PET/CT as the molecular multimodality imaging platform of choice," *Medical Physics*, vol. 34, no. 5, pp. 1525–1528, 2007.
- [46] W.-D. Heiss, "The potential of PET/MR for brain imaging," European Journal of Nuclear Medicine and Molecular Imaging, vol. 36, no. 1, pp. 105– 112, 2009.
- [47] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," Annual Review of Biomedical Engineering, vol. 2, no. 1, pp. 315–337, 2000.
- [48] S. Hu, E. Hoffman, and J. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images," *IEEE Transactions* on Medical Imaging, vol. 20, no. 6, pp. 490–498, 2001.
- [49] W. Jentzen, L. Freudenberg, E. G. Eising, M. Heinze, W. Brandau, and A. Bockisch, "Segmentation of PET volumes by iterative image thresholding," *Journal of Nuclear Medicine*, vol. 48, no. 1, pp. 108–114, 2007.
- [50] R. Adams and L. Bischof, "Seeded region growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641–647, 1994.
- [51] J. Canny, "A computational approach to edge detection," *IEEE Transac*tions on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679– 698, 1986.
- [52] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, and T.-J. Chen, "Fuzzy cmeans clustering with spatial information for image segmentation," *Computerized Medical Imaging and Graphics*, vol. 30, no. 1, pp. 9–15, 2006.
- [53] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [54] M. Leventon, W. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 316–323, 2000.
- [55] B. Zitová and J. Flusser, "Image registration methods: a survey," Image and Vision Computing, vol. 21, no. 11, pp. 977–1000, 2003.
- [56] W. R. Crum, T. Hartkens, and D. L. G. Hill, "Non-rigid image registration: theory and practice," *British Journal of Radiology*, vol. 77 (Supplement 2), pp. S140–S153, 2004.

- [57] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "elastix: a toolbox for intensity-based medical image registration.," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, 2010.
- [58] H. Pfister, B. Lorensen, C. Bajaj, G. Kindlmann, W. Schroeder, L. Avila, K. Raghu, R. Machiraju, and J. Lee, "The transfer function bake-off," *IEEE Computer Graphics and Applications*, vol. 21, no. 3, pp. 16–22, 2001.
- [59] G. Kindlmann, R. Whitaker, T. Tasdizen, and T. Moller, "Curvature-based transfer functions for direct volume rendering: methods and applications," in *IEEE Visualization*, pp. 513–520, 2003.
- [60] L. Mroz, A. König, and E. Gröller, "Maximum intensity projection at warp speed," Computers & Graphics, vol. 24, no. 3, pp. 343 – 352, 2000.
- [61] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 2, no. 1, pp. 1–19, 2006.
- [62] Y. Rui, T. S. Huang, and S.-F. Chang, "Image retrieval: Current techniques, promising directions, and open issues," *Journal of Visual Communication and Image Representation*, vol. 10, no. 1, pp. 39–62, 1999.
- [63] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," ACM Computing Surveys, vol. 40, no. 2, pp. 5:1–5:60, 2008.
- [64] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content: the QBIC system," *Computer*, vol. 28, no. 9, pp. 23–32, 1995.
- [65] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. C. Jain, and C.-F. Shu, "Virage image search engine: an open framework for image management," in *Proceedings of SPIE 2670* (I. K. Sethi and R. C. Jain, eds.), pp. 76–87, 1996.
- [66] A. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: Content-based manipulation of image databases," *International Journal of Computer Vision*, vol. 18, no. 3, pp. 233–254, 1996.
- [67] G. Chechik, V. Sharma, U. Shalit, and S. Bengio, "Large scale online learning of image similarity through ranking," *Journal of Machine Learning Research*, vol. 11, pp. 1109–1135, 2010.
- [68] J. S. Duncan and N. Ayache, "Medical image analysis: progress over two decades and the challenges ahead," *IEEE Transactions on Pattern Analysis* and Machine Intelligence, vol. 22, no. 1, pp. 85–106, 2000.

- [69] A. Ferman, A. Tekalp, and R. Mehrotra, "Robust color histogram descriptors for video segment retrieval and identification," *IEEE Transactions on Image Processing*, vol. 11, no. 5, pp. 497–508, 2002.
- [70] A. Vailaya, M. Figueiredo, A. Jain, and H.-J. Zhang, "Image classification for content-based indexing," *IEEE Transactions on Image Processing*, vol. 10, no. 1, pp. 117–130, 2001.
- [71] K. van de Sande, T. Gevers, and C. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Transactions on Pattern Analysis* and Machine Intelligence, vol. 32, no. 9, pp. 1582–1596, 2010.
- [72] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.
- [73] L. Tesař, A. Shimizu, D. Smutek, H. Kobatake, and S. Nawano, "Medical image analysis of 3D CT images based on extension of Haralick texture features," *Computerized Medical Imaging and Graphics*, vol. 32, no. 6, pp. 513–520, 2008.
- [74] B. Manjunath and W. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelli*gence, vol. 18, no. 8, pp. 837–842, 1996.
- [75] M. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Transactions on Image Processing*, vol. 11, no. 2, pp. 146–158, 2002.
- [76] M.-K. Hu, "Visual pattern recognition by moment invariants," IRE Transactions on Information Theory, vol. 8, no. 2, pp. 179–187, 1962.
- [77] T. Sebastian, P. Klein, and B. Kimia, "Recognition of shapes by editing their shock graphs," *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, vol. 26, pp. 550–571, may 2004.
- [78] S. Berretti, A. D. Bimbo, and P. Pala, "Description, matching and retrieval by content of 3D objects," in *Digital Libraries: Research and Development*, vol. 4877 of *Lecture Notes in Computer Science*, pp. 288–297, Springer-Verlag Berlin, Heidelberg, 2007.
- [79] J. W. H. Tangelder and R. C. Veltkamp, "A survey of content based 3D shape retrieval methods," *Multimedia Tools and Applications*, vol. 39, no. 3, pp. 441–471, 2008.
- [80] X. Zhou, C. Ang, and T. Ling, "Image retrieval based on object's orientation spatial relationship," *Pattern Recognition Letters*, vol. 22, no. 5, pp. 469– 477, 2001.

- [81] B. Fischer, C. J. Thies, M. O. Guld, and T. M. Lehmann, "Contentbased image retrieval by matching hierarchical attributed region adjacency graphs," in *Proceedings of SPIE 5370* (J. M. Fitzpatrick and M. Sonka, eds.), pp. 598–606, 2004.
- [82] N. Alajlan, M. Kamel, and G. Freeman, "Geometry-based image retrieval in binary image databases," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 1003–1013, 2008.
- [83] M. Huang, W. Yang, M. Yu, Z. Lu, Q. Feng, and W. Chen, "Retrieval of brain tumors with region-specific bag-of-visual-words representations in contrast-enhanced MRI images," *Computational and Mathematical Methods* in *Medicine*, vol. 2012, p. 17, 2012.
- [84] N. Hoàng, V. Gouet-Brunet, M. Rukoz, and M. Manouvrier, "Embedding spatial information into image content description for scene retrieval," *Pattern Recognition*, vol. 43, no. 9, pp. 3013–3024, 2010.
- [85] T. Jaworska, J. Kacprzyk, N. Marín, and S. Zadrożny, "On dealing with imprecise information in a content based image retrieval system," in *Computational Intelligence for Knowledge-Based Systems Design* (E. Hüllermeier, R. Kruse, and F. Hoffmann, eds.), vol. 6178 of *Lecture Notes in Computer Science*, pp. 149–158, Springer Berlin Heidelberg, 2010.
- [86] S. Wang, D. Liu, and H. L. Fangming Gu, Feng Yang, "Similar matching for images with complex spatial relations," *Journal of Computational Information Systems*, vol. 8, no. 21, pp. 8727–8734, 2012.
- [87] S. Aksoy and R. M. Haralick, "Feature normalization and likelihoodbased similarity measures for image retrieval," *Pattern Recognition Letters*, vol. 22, no. 5, pp. 563–582, 2001.
- [88] P. W. Kwan, K. Kameyama, and K. Toraichi, "On a relaxation-labeling algorithm for real-time contour-based image similarity retrieval," *Image* and Vision Computing, vol. 21, no. 3, pp. 285–294, 2003.
- [89] K. Shearer, H. Bunke, and S. Venkatesh, "Video indexing and similarity retrieval by largest common subgraph detection using decision trees," *Pattern Recognition*, vol. 34, no. 5, pp. 1075–1091, 2001.
- [90] S. Berretti, A. Del Bimbo, and E. Vicario, "Efficient matching and indexing of graph models in content-based retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1089–1105, 2001.
- [91] W. Cai, J. Kim, and D. Feng, "Content-based medical image retrieval," in Biomedical Information Technology (D. D. Feng, ed.), pp. 83–114, Elsevier, 2008.

- [92] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood, "Does organisation by similarity assist image browsing?," in *Proceedings of the SIGCHI Confer*ence on Human Factors in Computing Systems, pp. 190–197, 2001.
- [93] R. W. White and R. A. Roth, "Exploratory search: Beyond the queryresponse paradigm," Synthesis Lectures on Information Concepts, Retrieval, and Services, vol. 1, pp. 1–98, 2009.
- [94] D. Heesch, "A survey of browsing models for content based image retrieval," Multimedia Tools and Applications, vol. 40, no. 2, pp. 261–284, 2008.
- [95] M. L. Wilson, "Search user interface design," Synthesis Lectures on Information Concepts, Retrieval, and Services, vol. 3, no. 3, pp. 1–143, 2011.
- [96] M. Tory and T. Moller, "Human factors in visualization research," *IEEE Transactions on Visualization and Computer Graphics*, vol. 10, no. 1, pp. 72–84, 2004.
- [97] X. S. Zhou and T. S. Huang, "Relevance feedback in image retrieval: A comprehensive review," *Multimedia Systems*, vol. 8, pp. 536–544, 2003.
- [98] R. Haux, "Health information systems past, present, future," International Journal of Medical Informatics, vol. 75, no. 3-4, pp. 268 – 281, 2006.
- [99] A. Holt, I. Bichindaritz, R. Schmidt, and P. Perner, "Medical applications in case-based reasoning," *The Knowledge Engineering Review*, vol. 20, no. 3, pp. 289–292, 2005.
- [100] C.-R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick, "ASSERT: A physician-in-the-loop content-based retrieval system for HRCT image databases," *Computer Vision and Image Understand-ing*, vol. 75, no. 1-2, pp. 111–132, 1999.
- [101] L. R. Long, S. Antani, T. M. Deserno, and G. R. Thoma, "Content-based image retrieval in medicine: Retrospective assessment, state of the art, and future directions.," *International Journal of Healthcare Information Systems and Informatics*, vol. 4, no. 1, pp. 1–16, 2009.
- [102] C. Akgül, D. Rubin, S. Napel, C. Beaulieu, H. Greenspan, and B. Acar, "Content-based image retrieval in radiology: Current status and future directions," *Journal of Digital Imaging*, vol. 24, no. 2, pp. 208–222, 2011.
- [103] D. Keysers, J. Dahmen, H. Ney, B. B. Wein, and T. M. Lehmann, "Statistical framework for model-based image retrieval in medical applications," *Journal of Electronic Imaging*, vol. 12, no. 1, pp. 59–68, 2003.
- [104] M. O. Güld, C. Thies, B. Fischer, and T. M. Lehmann, "A generic concept for the implementation of medical image retrieval systems," *International Journal of Medical Informatics*, vol. 76, no. 2-3, pp. 252–259, 2007.

- [105] D. Iakovidis, N. Pelekis, E. Kotsifakos, I. Kopanakis, H. Karanikas, and Y. Theodoridis, "A pattern similarity scheme for medical image retrieval," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 442–450, 2009.
- [106] S. Antani, D. Lee, L. R. Long, and G. R. Thoma, "Evaluation of shape similarity measurement methods for spine X-ray images," *Journal of Visual Communication and Image Representation*, vol. 15, no. 3, pp. 285–302, 2004.
- [107] S. Antani, L. R. Long, G. R. Thoma, and D.-J. Lee, "Evaluation of shape indexing methods for content-based retrieval of x-ray images," in *Proceed*ings of SPIE 5021 (M. M. Yeung, R. W. Lienhart, and C.-S. Li, eds.), pp. 405–416, 2003.
- [108] D.-J. Lee, S. Antani, and L. R. Long, "Similarity measurement using polygon curve representation and fourier descriptors for shape-based vertebral image retrieval," in *Proceedings of SPIE 5032* (M. Sonka and J. M. Fitzpatrick, eds.), pp. 1283–1291, 2003.
- [109] X. Xu, D.-J. Lee, S. Antani, and L. Long, "A spine x-ray image retrieval system using partial shape matching," *Information Technology in Biomedicine*, *IEEE Transactions on*, vol. 12, pp. 100–108, jan. 2008.
- [110] W. Hsu, S. Antani, L. R. Long, L. Neve, and G. R. Thoma, "SPIRS: A webbased image retrieval system for large biomedical databases," *International Journal of Medical Informatics*, vol. 78 (Supplement 1), pp. S13–S24, 2009.
- [111] D.-J. Lee, S. Antani, Y. Chang, K. Gledhill, L. R. Long, and P. Christensen, "CBIR of spine X-ray images on inter-vertebral disc space and shape profiles using feature ranking and voting consensus," *Data & Knowledge Engineer*ing, vol. 68, no. 12, pp. 1359–1369, 2009.
- [112] X. Qian, H. D. Tagare, R. K. Fulbright, R. Long, and S. Antani, "Optimal embedding for shape indexing in medical image databases," *Medical Image Analysis*, vol. 14, no. 3, pp. 243 – 254, 2010.
- [113] Z. Xue, S. Antani, L. R. Long, J. Jeronimo, and G. R. Thoma, "Investigating CBIR techniques for cervicographic images.," in *Proceedings of the Annual Symposium of American Medical Information Association*, pp. 826– 830, 2007.
- [114] P. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, and Z. Protopapas, "Fast and effective retrieval of medical tumor shapes," *IEEE Transactions on Knowledge and Data Engineering*, vol. 10, no. 6, pp. 889–904, 1998.
- [115] L. Yang, R. Jin, L. Mummert, R. Sukthankar, A. Goode, B. Zheng, S. Hoi, and M. Satyanarayanan, "A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval,"

IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 1, pp. 30–44, 2010.

- [116] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, and C. Roux, "Wavelet optimization for content-based image retrieval in medical databases," *Medical Image Analysis*, vol. 14, no. 2, pp. 227 – 241, 2010.
- [117] G. Quellec, M. Lamard, L. Bekri, G. Cazuguel, C. Roux, and B. Cochener, "Medical case retrieval from a committee of decision trees," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1227– 1235, 2010.
- [118] G. Quellec, M. Lamard, G. Cazuguel, C. Roux, and B. Cochener, "Case retrieval in medical databases by fusing heterogeneous information," *IEEE Transactions on Medical Imaging*, vol. 30, no. 1, pp. 108–118, 2011.
- [119] J. G. Dy, C. E. Brodley, A. Kak, L. S. Broderick, and A. M. Aisen, "Unsupervised feature selection applied to content-based retrieval of lung images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 3, pp. 373–378, 2003.
- [120] D. Unay, A. Ekin, and R. Jasinschi, "Local structure-based region-ofinterest retrieval in brain MR images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 4, pp. 897–903, 2010.
- [121] E. G. Petrakis, "Design and evaluation of spatial similarity approaches for image retrieval," *Image and Vision Computing*, vol. 20, no. 1, pp. 59–76, 2002.
- [122] W. Cai, D. Feng, and R. Fulton, "Content-based retrieval of dynamic PET functional images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 4, no. 2, pp. 152–158, 2000.
- [123] J. Kim, W. Cai, D. Feng, and H. Wu, "A new way for multidimensional medical data management: Volume of interest (VOI)-based retrieval of medical images with visual and functional features," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 3, pp. 598–607, 2006.
- [124] J. Kim, L. Constantinescu, W. Cai, and D. D. Feng, "Content-based dualmodality biomedical data retrieval using co-aligned functional and anatomical features," in *Proceedings of the MICCAI Workshop on Content-based Image Retrieval for Biomedical Image Archives: achievements, problems* and prospects, pp. 45–52, 2007.
- [125] Y. Song, W. Cai, S. Eberl, M. Fulham, and D. Feng, "A content-based image retrieval framework for multi-modality lung images," in *Proceedings* of the 23rd International Symposium on Computer-Based Medical Systems, pp. 285–290, 2010.

- [126] Y. Song, W. Cai, S. Eberl, M. Fulham, and D. Feng, "Structure-adaptive feature extraction and representation for multi-modality lung images retrieval," in *International Conference on Digital Image Computing: Techniques and Applications*, pp. 152–157, 2010.
- [127] Y. Song, W. Cai, S. Eberl, M. Fulham, and D. Feng, "Thoracic image case retrieval with spatial and contextual information," in *IEEE International* Symposium on Biomedical Imaging: From Nano to Macro,, pp. 1885–1888, 2011.
- [128] Y. Song, W. Cai, S. Eberl, M. Fulham, and D. Feng, "Thoracic image matching with appearance and spatial distribution," in *Proceedings of the* 33rd International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4469–4472, 2011.
- [129] Y. Song, W. Cai, and D. Feng, "Hierarchical spatial matching for medical image retrieval," in *Proceedings of the International ACM workshop on Medical Multimedia Analysis and Retrieval*, pp. 1–6, 2011.
- [130] W. Cai, Y. Song, and D. D. Feng, "Regression and classification based distance metric learning for medical image retrieval," in *IEEE International* Symposium on Biomedical Imaging, pp. 1775–1778, 2012.
- [131] A. Kumar, J. Kim, W. Cai, S. Eberl, and D. Feng, "A graph-based approach to the retrieval of dual-modality biomedical images using spatial relationships," in *Proceedings of the 30th International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 390–393, 2008.
- [132] A. Kumar, J. Kim, S. Eberl, M. Fulham, and D. Feng, "Content-based retrieval of PET-CT lung cancer data using spatial features," *Journal of Nuclear Medicine*, vol. 50 (Supplement 2), p. 176, 2009.
- [133] A. Kumar, J. Kim, L. Wen, and D. Feng, "A graph-based approach to the retrieval of volumetric PET-CT lung images," in *Proceedings of the 34th International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5408–5411, 2012.
- [134] A. Kumar, J. Kim, D. Feng, and M. Fulham, "Graph-based retrieval of multi-modality medical images: A comparison of representations using simulated images," in *Proceedings of the 25th International Symposium on Computer-Based Medical Systems*, pp. 1–6, 2012.
- [135] A. Kumar, D. Haraguchi, J. Kim, L. Wen, S. Eberl, M. Fulham, and D. Feng, "A query and visualisation interface for a PET-CT image retrieval system," *International Journal of Computer Assisted Radiology and Surgery*, vol. 6 (Supplement 1), pp. 69–69, 2011.

- [136] A. Kumar, J.Kim, L. Bi, , and D. Feng, "An image retrieval interface for volumetric multi-modal medical data: application to PET-CT contentbased image retrieval," *International Journal of Computer Assisted Radiology and Surgery*, vol. 7 (Supplement 1), pp. 475–477, 2012.
- [137] A. Kumar, J. Kim, L. Bi, M. Fulham, and D. Feng, "Designing user interfaces to enhance human interpretation of medical CBIR: application to PET-CT images," *International Journal of Computer Assisted Radiology* and Surgery, vol. 8 (Supplement 1), pp. S72–73, 2013.
- [138] A. Kumar, J. Kim, L. Wen, M. Fulham, and D. Feng, "A graph-based approach for the retrieval of multi-modality medical images," *Medical Image Analysis*, under review (undergoing revisions).
- [139] A. Kumar, J. Kim, L. Bi, M. Fulham, and D. Feng, "Designing user interfaces to enhance human interpretation of medical content-based image retrieval: application to PET-CT images," *International Journal of Computer Assisted Radiology and Surgery*, 2013. EPub May 2013. doi: 10.1007/s11548-013-0896-5.
- [140] A. Kumar, J. Kim, and D. Feng, "Graph-based retrieval of PET-CT images using vector space embedding," in *Proceedings of the 26th International* Symposium on Computer-Based Medical Systems. In press.
- [141] S. Radhouani, J. Lim, J.-P. Chevallet, and G. Falquet, "Combining textual and visual ontologies to solve medical multimodal queries," in *IEEE International Conference on Multimedia and Expo*, pp. 1853–1856, 2006.
- [142] C. Lacoste, J.-H. Lim, J.-P. Chevallet, and D. Le, "Medical-image retrieval based on knowledge-assisted text and image indexing," *IEEE Transactions* on Circuits and Systems for Video Technology, vol. 17, no. 7, pp. 889–900, 2007.
- [143] J. Gobeill, H. Müller, and P. Ruch, "Translation by text categorisation: Medical image retrieval in imageclefmed 2006," in *Evaluation of Multilingual and Multi-modal Information Retrieval* (C. Peters, P. Clough, F. Gey, J. Karlgren, B. Magnini, D. Oard, M. de Rijke, and M. Stempfhuber, eds.), vol. 4730 of *Lecture Notes in Computer Science*, pp. 706–710, Springer Berlin / Heidelberg, 2007.
- [144] J. Villena-Román, S. Lana-Serrano, and J. González-Cristóbal, "MIRACLE at ImageCLEFmed 2007: Merging textual and visual strategies to improve medical image retrieval," in Advances in Multilingual and Multimodal Information Retrieval (C. Peters, V. Jijkoun, T. Mandl, H. Müller, D. Oard, A. Peñas, V. Petras, and D. Santos, eds.), vol. 5152 of Lecture Notes in Computer Science, pp. 593–596, Springer Berlin / Heidelberg, 2008.
- [145] J. C. Caicedo, J. G. Moreno, E. A. Niño, and F. A. González, "Combining visual features and text data for medical image retrieval using latent

semantic kernels," in *Proceedings of the International Conference on Mul*timedia Information Retrieval, pp. 359–366, 2010.

- [146] M. Rahman, S. Antani, R. Long, D. Demner-Fushman, and G. Thoma, "Multi-modal query expansion based on local analysis for medical image retrieval," in *Medical Content-Based Retrieval for Clinical Decision Support* (B. Caputo, H. Müller, T. Syeda-Mahmood, J. Duncan, F. Wang, and J. Kalpathy-Cramer, eds.), vol. 5853 of *Lecture Notes in Computer Science*, pp. 110–119, Springer Berlin / Heidelberg, 2010.
- [147] H. Müller, J. Kalpathy-Cramer, J. Charles E. Kahn, and W. Hersh, "Comparing the quality of accessing medical literature using content-based visual and textual information retrieval," in *Proceedings of SPIE 7264* (K. M. Siddiqui and B. J. Liu, eds.), pp. 726405:1–11, 2009.
- [148] W. W. Chu, I. T. Ieong, and R. K. Taira, "A semantic modeling approach for image retrieval by content," *The VLDB Journal - The International Journal on Very Large Data Bases*, vol. 3, no. 4, pp. 445–477, 1994.
- [149] W. Chu, C.-C. Hsu, A. Cardenas, and R. Taira, "Knowledge-based image retrieval with spatial and temporal constructs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 10, no. 6, pp. 872–888, 1998.
- [150] A. Névéol, T. M. Deserno, S. J. Darmoni, M. O. Güld, and A. R. Aronson, "Natural language processing versus content-based image analysis for medical document retrieval," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 1, pp. 123–134, 2009.
- [151] C. P. Langlotz, "RadLex: A new method for indexing online educational materials," *Radiographics*, vol. 26, no. 6, pp. 1595–1597, 2006.
- [152] H. Müller, T. Deselaers, T. Deserno, J. Kalpathy-Cramer, E. Kim, and W. Hersh, "Overview of the ImageCLEFmed 2007 medical retrieval and medical annotation tasks," in Advances in Multilingual and Multimodal Information Retrieval (C. Peters, V. Jijkoun, T. Mandl, H. Müller, D. Oard, A. Peñas, V. Petras, and D. Santos, eds.), vol. 5152 of Lecture Notes in Computer Science, pp. 472–491, Springer Berlin / Heidelberg, 2008.
- [153] H. Müller, J. Kalpathy-Cramer, C. Kahn, W. Hatt, S. Bedrick, and W. Hersh, "Overview of the ImageCLEFmed 2008 medical image retrieval task," in *Evaluating Systems for Multilingual and Multimodal Information* Access (C. Peters, T. Deselaers, N. Ferro, J. Gonzalo, G. Jones, M. Kurimo, T. Mandl, A. Peñas, and V. Petras, eds.), vol. 5706 of Lecture Notes in Computer Science, pp. 512–522, Springer Berlin / Heidelberg, 2009.
- [154] H. Müller, J. Kalpathy-Cramer, I. Eggel, S. Bedrick, S. Radhouani, B. Bakke, C. Kahn, and W. Hersh, "Overview of the CLEF 2009 medical image retrieval track," in *Multilingual Information Access Evaluation II. Multimedia Experiments* (C. Peters, B. Caputo, J. Gonzalo, G. Jones,

J. Kalpathy-Cramer, H. Müller, and T. Tsikrika, eds.), vol. 6242 of *Lecture* Notes in Computer Science, pp. 72–84, Springer Berlin / Heidelberg, 2010.

- [155] J. Liu, Y. Hu, M. Li, S. Ma, and W. ying Ma, "Medical image annotation and retrieval using visual features," in *Evaluation of Multilingual and Multi-modal Information Retrieval*, vol. 4730 of *Lecture Notes in Computer Science*, pp. 678–685, Springer-Verlag Berlin, Heidelberg, 2007.
- [156] M. M. Rahman, B. C. Desai, and P. Bhattacharya, "Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion," *Computerized Medical Imaging and Graphics*, vol. 32, no. 2, pp. 95 – 108, 2008.
- [157] H. Akakin and M. Gurcan, "Content-based microscopic image retrieval system for multi-image queries," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 4, pp. 758–769, 2012.
- [158] G. Allampalli-Nagaraj and I. Bichindaritz, "Automatic semantic indexing of medical images using a web ontology language for case-based image retrieval," *Engineering Applications of Artificial Intelligence*, vol. 22, no. 1, pp. 18–25, 2009.
- [159] X. Zhou, R. Stern, and H. Müller, "Case-based fracture image retrieval," International Journal of Computer Assisted Radiology and Surgery, vol. 7, no. 3, pp. 401–411, 2012.
- [160] S. C. Huang, M. E. Phelps, E. J. Hoffman, K. Sideris, C. J. Selin, and D. E. Kuhl, "Noninvasive determination of local cerebral metabolic rate of glucose in man," *American Journal of Physiology - Endocrinology And Metabolism*, vol. 238, no. 1, pp. E69–E82, 1980.
- [161] E. Chang, K. Goh, G. Sychay, and G. Wu, "CBSA: content-based soft annotation for multimodal image retrieval using Bayes point machines," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 1, pp. 26–38, 2003.
- [162] W. Hersh, H. Müller, and J. Kalpathy-Cramer, "The ImageCLEFmed medical image retrieval task test collection," *Journal of Digital Imaging*, vol. 22, no. 6, pp. 648–655, 2009.
- [163] G. Tian, H. Fu, and D. Feng, "Automatic medical image categorization and annotation using LBP and MPEG-7 edge histograms," in *International Conference on Information Technology and Applications in Biomedicine*, pp. 51–53, 2008.
- [164] V. Spitzer, M. J. Ackerman, A. L. Scherzinger, and D. Whitlock, "The visible human male: A technical report," *Journal of the American Medical Informatics Association*, vol. 3, no. 2, pp. 118–130, 1996.

- [165] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.
- [166] G. Langs, H. Müller, B. H. Menze, and A. Hanbury, "VISCERAL: Towards large data in medical imaging - challenges and directions," in *Medical Content-Based Retrieval for Clinical Decision Support* (H. Greenspan, H. Müller, and T. Syeda-Mahmood, eds.), vol. 7723 of *Lecture Notes in Computer Science*, pp. 92–98, 2013.
- [167] A. Ekin, R. Jasinschi, J. van der Grond, and M. Van Buchem, "Improving information quality of MR brain images by fully automatic and robust image analysis methods," *Journal of the Society for Information Display*, vol. 15, no. 6, pp. 367–376, 2007.
- [168] E. M. van Rikxoort, I. Isgum, Y. Arzhaeva, M. Staring, S. Klein, M. A. Viergever, J. P. Pluim, and B. van Ginneken, "Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus," *Medical Image Analysis*, vol. 14, no. 1, pp. 39–49, 2010.
- [169] Y. Jing, H. Rowley, C. Rosenberg, J. Wang, M. Zhao, and M. Covell, "Google image swirl, a large-scale content-based image browsing system," in *IEEE International Conference on Multimedia and Expo*, p. 267, 2010.
- [170] J. Etzold, A. Brousseau, P. Grimm, and T. Steiner, "Context-aware querying for multimodal search engines," in *Advances in Multimedia Modeling* (K. Schoeffmann, B. Merialdo, A. Hauptmann, C.-W. Ngo, Y. Andreopoulos, and C. Breiteneder, eds.), vol. 7131 of *Lecture Notes in Computer Science*, pp. 728–739, Springer Berlin / Heidelberg, 2012.
- [171] F. C. Detterbeck, D. J. Boffa, and L. T. Tanoue, "The new lung cancer staging system," *Chest*, vol. 136, no. 1, pp. 260–271, 2009.
- [172] S. B. Edge, D. R. Byrd, C. C. Compton, A. G. Frtiz, F. L. Greene, and A. Trotti, eds., AJCC Cancer Staging Manual. Springer New York, 2010.
- [173] O. Bodenreider, "The unified medical language system (UMLS): integrating biomedical terminology," *Nucleic Acids Research*, vol. 32 (Supplement 1), pp. D267–D270, 2004.
- [174] S. G. Armato III, G. McLennan, M. F. McNitt-Gray, C. R. Meyer, D. Yankelevitz, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, A. P. Reeves, B. Y. Croft, L. P. Clarke, and F. the Lung Image Database Consortium Research Group, "Lung image database consortium: Developing a resource for the medical imaging research community," *Radiology*, vol. 232, no. 3, pp. 739–748, 2004.
- [175] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies (OASIS): Crosssectional MRI data in young, middle aged, nondemented, and demented

older adults," *Journal of Cognitive Neuroscience*, vol. 19, no. 9, pp. 1498–1507, 2007.

- [176] H. Bunke and K. Riesen, "Towards the unification of structural and statistical pattern recognition," *Pattern Recognition Letters*, vol. 33, no. 7, pp. 811–825, 2012.
- [177] E. Tanaka, "Theoretical aspects of syntactic pattern recognition," Pattern Recognition, vol. 28, no. 7, pp. 1053–1061, 1995.
- [178] A. Jain, R. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.
- [179] H. Bunke and K. Riesen, "Recent advances in graph-based pattern recognition with applications in document analysis," *Pattern Recognition*, vol. 44, no. 5, pp. 1057–1067, 2011.
- [180] K. S. Fu, "A step towards unification of syntactic and statistical pattern recognition," *IEEE Transactions on Pattern Analysis and Machine Intelli*gence, vol. 8, no. 3, pp. 398–404, 1986.
- [181] J. Llados, E. Marti, and J. Villanueva, "Symbol recognition by errortolerant subgraph matching between region adjacency graphs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1137–1143, 2001.
- [182] H. Bunke, "Error correcting graph matching: on the influence of the underlying cost function," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 917–922, 1999.
- [183] E. Petrakis, C. Faloutsos, and K.-I. Lin, "ImageMap: an image indexing method based on spatial similarity," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 5, pp. 979–987, 2002.
- [184] A. Torsello, D. Hidovic-Rowe, and M. Pelillo, "Polynomial-time metrics for attributed trees," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 7, pp. 1087–1099, 2005.
- [185] W. Christmas, J. Kittler, and M. Petrou, "Structural matching in computer vision using probabilistic relaxation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 749–764, 1995.
- [186] P. Hart, N. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, no. 2, pp. 100–107, 1968.
- [187] M. Neuhaus, K. Riesen, and H. Bunke, "Fast suboptimal algorithms for the computation of graph edit distance," in *Structural, Syntactic, and Statistical Pattern Recognition* (D.-Y. Yeung, J. Kwok, A. Fred, F. Roli,

and D. de Ridder, eds.), vol. 4109 of *Lecture Notes in Computer Science*, pp. 163–172, Springer Berlin / Heidelberg, 2006.

- [188] J. Munkres, "Algorithms for the assignment and transportation problems," Journal of the Society for Industrial and Applied Mathematics, vol. 5, no. 1, pp. 32–38, 1957.
- [189] F. Bourgeois and J.-C. Lassalle, "An extension of the Munkres algorithm for the assignment problem to rectangular matrices," *Communications of the ACM*, vol. 14, no. 12, pp. 802–804, 1971.
- [190] B. Scholkopf and A. J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. Cambridge, MA, USA: MIT Press, 2002.
- [191] J. Shawe-Taylor and N. Cristianini, Kernel Methods for Pattern Analysis. New York, NY, USA: Cambridge University Press, 2004.
- [192] M. Neuhaus and H. Bunke, Bridging the Gap Between Graph Edit Distance and Kernel Machines. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 2007.
- [193] T. Gärtner, P. Flach, and S. Wrobel, "On graph kernels: Hardness results and efficient alternatives," in *Learning Theory and Kernel Machines* (B. Schölkopf and M. Warmuth, eds.), vol. 2777 of *Lecture Notes in Computer Science*, pp. 129–143, Springer Berlin / Heidelberg, 2003.
- [194] M. Neuhaus, K. Riesen, and H. Bunke, "Novel kernels for error-tolerant graph classification," *Spatial Vision*, vol. 22, no. 5, pp. 425–441, 2009.
- [195] P. P. Carbone, H. S. Kaplan, K. Musshoff, D. W. Smithers, and M. Tubiana, "Report of the committee on Hodgkin's disease staging classification," *Cancer Research*, vol. 31, no. 11, pp. 1860–1861, 1971.
- [196] L. Constantinescu, J. Kim, A. Kumar, D. Haraguchi, L. Wen, and D. Feng, "A patient-centric distribution architecture for medical image sharing," *Health Information Science and Systems*, vol. 1, article 3, 2013.
- [197] H. Wadell, "Volume, Shape, and Roundness of Quartz Particles," Journal of Geology, vol. 43, no. 3, pp. 250–280, 1935.
- [198] S. Tomei, A. Reilhac, D. Visvikis, N. Boussion, C. Odet, F. Giammarile, and C. Lartizien, "OncoPET_DB: A freely distributed database of realistic simulated whole body 18F-FDG PET images for oncology," *IEEE Transactions on Nuclear Science*, vol. 57, no. 1, pp. 246–255, 2010.
- [199] J. Bradley, W. L. Thorstad, S. Mutic, T. R. Miller, F. Dehdashti, B. A. Siegel, W. Bosch, and R. J. Bertrand, "Impact of FDG-PET on radiation therapy volume delineation in non-small-cell lung cancer.," *International Journal of Radiation Oncology*Biology*Physics*, vol. 59, no. 1, pp. 78–86, 2004.

- [200] M. D. Abramoff, P. J. Magelhaes, and S. J. Ram, "Image processing with ImageJ," *Biophotonics International*, vol. 11, no. 7, pp. 36–42, 2004.
- [201] J. O'Madadhain, D. Fisher, S. White, and Y. Boey, "The JUNG (Java Universal Network/Graph) framework," http://jung.sourceforge.net/, First Published: 01/08/2003, Last Updated: 24/01/2010, Last Checked: 25/03/2013.
- [202] J. Nielsen and T. K. Landauer, "A mathematical model of the finding of usability problems," in *Proceedings of the INTERACT '93 and CHI '93* conference on Human factors in computing systems, pp. 206–213, 1993.
- [203] E. L.-C. Law, V. Roto, M. Hassenzahl, A. P. Vermeeren, and J. Kort, "Understanding, scoping and defining user experience: a survey approach," in *Proceedings of the SIGCHI Conference on Human Factors in Computing* Systems, pp. 719–728, 2009.
- [204] Y. Zhan, X. S. Zhou, Z. Peng, and A. Krishnan, "Active scheduling of organ detection and segmentation in whole-body medical images," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008* (D. Metaxas, L. Axel, G. Fichtinger, and G. Székely, eds.), vol. 5241 of *Lecture Notes in Computer Science*, pp. 313–321, Springer Berlin Heidelberg, 2008.
- [205] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," Journal of Machine Learning Research, vol. 3, pp. 1157–1182, 2003.
- [206] K. Riesen and H. Bunke, "Approximate graph edit distance computation by means of bipartite graph matching," *Image and Vision Computing*, vol. 27, no. 7, pp. 950–959, 2009.
- [207] K. Riesen and H. Bunke, "Graph classification based on vector space embedding," International Journal of Pattern Recognition and Artificial Intelligence, vol. 23, no. 6, pp. 1053–1081, 2009.
- [208] P. Zhu, C. Zhang, H. Li, R. C. C. Cheung, and B. Hu, "An FPGA-based acceleration platform for auction algorithm," in *IEEE International Sym*posium on Circuits and Systems, pp. 1002–1005, 2012.
- [209] M. Deshpande, M. Kuramochi, N. Wale, and G. Karypis, "Frequent substructure-based approaches for classifying chemical compounds," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 8, pp. 1036– 1050, 2005.
- [210] N. Max, "Optical models for direct volume rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 1, no. 2, pp. 99–108, 1995.
- [211] J. Thomas and K. Cook, "A visual analytics agenda," *IEEE Computer Graphics and Applications*, vol. 26, no. 1, pp. 10–13, 2006.