Computer-Aided Cancer Detection and Diagnosis
RECENT ADVANCES
Computer-Aided Cancer Detection and Diagnosis
RECENT ADVANCES

Jinshan Tang
Sos S. Agaian
Editors

SPIE PRESS
Bellingham, Washington  USA
# Table of Contents

*Preface* xi  
*List of Contributors* xvii  

## 1 Computer-Aided Detection of Colonic Polyps in CT Colonography 1

* Jianhua Yao  
  1.1 Colonic Polyps and Colon Cancer 1  
  1.2 CT Colonography 2  
  1.3 Computer-Aided Detection Using CTC 3  
    1.3.1 CAD pipeline 4  
    1.3.2 Colon segmentation 5  
    1.3.3 Supine–prone registration 6  
    1.3.4 Colon unfolding 8  
    1.3.5 Polyp segmentation 9  
    1.3.6 Polyp characterization and features 11  
    1.3.7 Machine learning and classification 12  
    1.3.8 Content-based image retrieval 13  
    1.3.9 CAD performance 14  
  1.4 Discussion 16  
  References 17  

## 2 Preprocessing Tools for Computer-Aided Cancer Imaging Systems 23

* Artyom M. Grigoryan and Sos S. Agaian  
  2.1 Introduction 24  
    2.1.1 Novel view on image processing 25  
  2.2 Transform-Based Image Enhancement 28  
    2.2.1 Quantitative measure of image enhancement 29  
  2.3 Tensor Representation of the Image 30  
  2.4 Decomposition by Direction Images 33  
  2.5 Tensor Transform Method of $\alpha$-Rooting 37  
    2.5.1 Effective formula for image enhancement 39  
    2.5.2 Algorithm of image enhancement by 1D $\alpha$-rooting 42  
  2.6 Decomposition by the 2D Paired Transform 42  
    2.6.1 Fourier transform splitting theorem 44  
    2.6.2 Complete set of the 2D paired transform 46
2.7 Paired Direction Images 47
   2.7.1 Principle of superposition by direction images 49
   2.7.2 Paired method of image enhancement 49
2.8 Enhancement by a Series of Direction Images 52
2.9 Compression: Multiresolution Map of the Image 55
   2.9.1 A-series linear transformation 56
2.10 Compression by the Tensor Transform 59
   2.10.1 Block-tensor-transform lossy image compression 62
2.11 Tensor Transform in Image Cryptography 65
2.12 Conclusion 69
References 70

3 Multimodality Imaging for Tumor Volume Definition in Radiation Oncology 79
   Issam El Naqa
   3.1 Introduction 79
   3.2 Single versus Multimodality Image Segmentation 80
   3.3 Methods for Multimodality Image Segmentation 82
      3.3.1 Multiple-image thresholding 83
      3.3.2 Clustering algorithms 83
         3.3.2.1 Fuzzy C-means algorithm 83
         3.3.2.2 Extending the fuzzy C-means algorithm to multiple images 84
      3.3.3 Active contour algorithms 85
         3.3.3.1 "Active-contour-without-edge" algorithm 87
         3.3.3.2 Extension to multiple images 88
   3.4 Examples of Multimodality Tumor Volume Definition 89
      3.4.1 PET/CT target definition in radiotherapy 89
      3.4.2 PET/CT segmentation of cervix cancer example 90
      3.4.3 MR/CT segmentation of prostate cancer example 91
      3.4.4 Coronary artery plaque MR image analysis 92
   3.5 Issues, Problems, and Future Directions 92
      3.5.1 Image understanding 93
      3.5.2 Deformable image registration 94
   3.6 Conclusions 94
References 94

4 Nonlinear Unsharp Masking for Enhancing Suspicious Regions in Mammograms 99
   Yicong Zhou, C. L. Philip Chen, Sos S. Agaian, and Karen Panetta
   4.1 Introduction 99
   4.2 Background 102
      4.2.1 Traditional unsharp masking 102
      4.2.2 The RUM algorithm 103
      4.2.3 The ANCE algorithm 104
6.2 Related Work

6.2.1 Related work on CE video analysis for automatic object detection

6.2.2 Related work on incremental learning using SVMs

6.3 Geometric Incremental Support Vector Machines

6.3.1 Geometric support vector machines

6.3.2 Geometric incremental support vector machine (GISVM)

6.4 Experimental Results and Discussion

6.4.1 Synthetic and benchmark data preparation

6.4.2 Parameter selection

6.4.3 Efficiency analysis

6.4.4 Accuracy analysis

6.4.5 Experiments with CE videos

6.5 Conclusion

References

7 Automated Melanoma Screening and Early Detection

Xiaojing Yuan, Ning Situ, Xiaohui Yuan, and George Zouridakis

7.1 Overview of Automated Melanoma Screening and Early Detection Systems

7.1.1 Optical imaging modalities for pigmented skin lesion image acquisition

7.1.2 Key functions of existing skin lesion classification systems

7.1.3 Review of representative skin cancer detection systems

7.1.4 Overview of skin lesions and commonly used criteria by clinicians

7.1.5 Pigmented skin lesion datasets

7.2 AutoScan: Automated Melanoma Screening and Early Detection

7.2.1 AutoScan preprocessing

7.2.2 AutoScan: region-of-interest identification

7.3 Mapping Computer-Generated Features to High-Level Concepts Used by Dermatologists

7.3.1 Computer-generated low-level features

7.3.2 Mapping high-level dermoscopic concepts with multiple-instance learning

7.3.2.1 Diverse density function and evidence confidence function

7.3.2.2 Boosting enhanced-instance prototype selection

7.3.3 Transforming a lesion image into a descriptor vector

7.3.4 Experiment setups and results

7.4 Integrating Feature Selection with Feature-Model Learning

7.4.1 Feature selection and combination

7.4.2 Multiple auxiliary kernel learning (MAKL): learning from heterogeneous feature spaces

7.4.3 Experiment setup and results
Table of Contents

7.5 Conclusions and Future Directions 201
References 201

8 A Complex Wavelet-Based Feature Extraction System for Microcalcification Detection in Digital Mammograms 211
Ping Zhang and Kwabena Agyepong
8.1 Introduction 211
8.2 System Design 213
8.3 Hybrid Feature Extraction 214
8.3.1 Surrounding region dependence-based method 214
8.3.2 Wavelet transform 216
8.3.3 Complex wavelet transform for feature extraction 217
8.3.4 2D-CWT multifractal feature 219
8.4 Classifier Design 220
8.5 Experiment Results 220
8.6 Conclusion 224
References 224

9 Computer-Aided Prostate Cancer Diagnosis: Principles, Recent Advances, and Future Prospective 229
Sos S. Agaian, Clara Mosquera-Lopez, Alejandro Velez-Hoyos, and Ian Thompson
9.1 Introduction 229
9.2 Clinical Approach for Prostate Cancer Detection and Grading 234
9.2.1 Core needle biopsy 234
9.2.2 Digital pathology imaging 235
9.3 State-of-the-Art in Histopathology-Image-Based, Computer-Aided Prostate Cancer Diagnosis 236
9.3.1 Image preprocessing 238
9.3.1.1 Color normalization 238
9.3.1.2 Histopathology image segmentation 239
9.3.2 Feature extraction 243
9.3.3 Classification 247
9.3.4 System accuracy assessment 248
9.3.4.1 Performance indicators 249
9.4 Conclusions, Future Directions, and Potential New Strategies 253
References 256

10 Analysis of Breast Masses in Mammograms Using the Fractal Dimension and Shape Factors 269
Grazia Raguso, Antonietta Ancona, Loredana Chieppa, Samuela L’Abbate, Maria Luisa Pepe, Fabio Mangieri, Shantanu Banik, and Rangaraj M. Rangayyan
10.1 Introduction 270
10.2 Methods 273
10.2.1 Fractal analysis 273
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.2.2</td>
<td>Shape factors</td>
</tr>
<tr>
<td>10.2.3</td>
<td>Feature analysis, selection, and classification</td>
</tr>
<tr>
<td>10.3</td>
<td>Datasets of Contours of Breast Masses</td>
</tr>
<tr>
<td>10.4</td>
<td>Results and Discussion</td>
</tr>
<tr>
<td>10.5</td>
<td>Conclusion</td>
</tr>
<tr>
<td>References</td>
<td></td>
</tr>
</tbody>
</table>

**11 Another Step towards Successful Tomographic Imaging in Cancer:**

**Solving the Problem of Image Reconstruction**

*Artyom M. Grigoryan*

11.1 Introduction 
11.1.1 CT images and lung cancer | 296
11.1.2 CT images and breast cancer | 296
11.1.3 Breast cancer with CT, mammography, and MRI | 297
11.1.4 Algorithms in CT | 297
11.2 Model of the Image | 298
11.2.1 Line integrals and ray sums | 299
11.2.2 Two types of parallel rays | 300
11.3 The Image and the Set of Splitting-Signals | 301
11.4 Geometry of the Projections on the Lattice | 305
11.4.1 Main equations for geometrical rays when $N$ is prime | 315
11.4.2 Simulation results for modeled images | 316
11.5 Geometry for the Lattice $N \times N$ when $N$ is a Power of Two | 318
11.5.1 Algorithm of image reconstruction | 326
11.5.2 Convolution equations | 327
11.5.3 Preliminary results | 329
11.6 Conclusion | 333
References | 334

*Index* | 339
Cancer is a significant threat to human life. Currently, the United States spends over 100 billion dollars annually on cancers, such as breast, lung, and prostate cancer. Based on statistics from the World Health Organization (WHO), deaths caused by cancer will reach approximately 12 million people in 2030. Thus, it has become a challenge to fight cancers in both medical practice and in the scientific research field. Imaging of cancer is an increasingly important component of understanding and treating cancer. Today, it is necessary to not only assess tumors morphologically but also provide information about the pathophysiological and metabolic aspects of tumor behavior with functional imaging techniques. Over the last two to three decades, the field of diagnostic cancer imaging has witnessed remarkable evolution that has affected virtually every aspect of research and clinical management of cancer. This evolution has been the result of innovations in three main aspects: innovative instrumentation (including a new class of scanners); development of new contrast agents and radiolabeled tracers; and imaging tools (including computer-aided detection or diagnosis technologies) for the detection, evolution, staging, and prognosis of many types of cancer. Current standard imaging techniques cannot accurately detect early diseases, and they provide limited information for disease staging.

The major goals of current cancer imaging are as follows:

- Provide more reliable disease characterization through the synthesis of anatomic, functional, and molecular imaging information;
- Refine and optimize imaging capabilities in oncology;
- Establish new imaging modalities and findings, and discover the potential use of these techniques;
- Find more individualized assessment of tumor biology, personalized treatments, and response to treatment;
- Develop image-processing-based cancer control systems; and
- Explore imaging capabilities and strategies to streamline cancer drug development.
Six levels of assessment to determine the efficacy of diagnostic imaging should be considered:

- Technical performance – the ability to obtain a high image quality.
- Diagnostic performance – the ability to identify a disease correctly.
- Diagnostic impact – a measure of sensitivity, specificity, positive predictive value, negative predictive value, and accuracy.
- Therapeutic impact – the influence an imaging result has on clinical diagnostic confidence.
- Alteration in management based on the results of imaging.
- Impact on health – the influence of imaging on the disease outcome.

Computer-aided detection or diagnosis (CAD) technologies play a key role in the detection of cancers and help reduce the death rate; as such, they have greatly advanced over the past decades. The aim of this book is to publish and promote high-quality research in key technologies used in computer-aided cancer detection and diagnosis systems. The following 11 chapters cover different types of cancers, including skin cancer, breast cancer, prostate cancer, colon cancer, etc.; they also span different scientific fields, such as biomedicine, imaging, image processing, pattern recognition, system analysis, etc.

Colonic polyps are fleshy growths that appear on the inside of the large intestine, and certain types of polyps grow large enough and can become cancerous. Screening for colon polyps and removing them before they become cancerous can reduce the risk of colon cancer. Chapter 1 reviews computer-aided systems and technologies for colonic polyp detection using CT colonography. It introduces the history, preparation, imaging protocol, and clinical value of CTC and related image processing technologies, including colon segmentation, supine–prone registration, colon unfolding, polyp segmentation and characterization, classification, and content-based image retrieval. It also summarizes the performances and limitations of various CAD systems.

Digital image processing technologies have important applications in computer-aided cancer imaging systems, and they play a key role in cancer detection. These technologies include image enhancement, image segmentation, image compression, image encryption, etc., all of which are needed to provide information about the extent of disease and help plan treatment of the cancer. However, these technologies are not fully developed, and further investigation is needed to improve the accuracy of computer-aided cancer detection systems. Chapter 2 examines three image processing technologies

(image enhancement, image compression, and image encryption) that are often adopted by computer cancer detection systems. Based on nontraditional representation of images in the form of 1D independent signals, new approaches for enhancement, compression, and encryption are presented as a preprocessing tool for computer-aided imaging systems.

Chapter 3 presents an overview of recent advances in multimodality imaging technologies for diagnostic radiology and image-guided radiotherapy. In particular, it discusses the expanding role of multimodality imaging in cancer detection and segmentation for radiation oncology. Using complementary information from multimodality images significantly improves the robustness and accuracy of tumor volume definitions in radiotherapeutic treatments of cancer. The chapter also provides working examples for developing algorithms for multimodality target volume definitions in different cancers and highlights the potential opportunities in this field for computer-aided detection and image-guided treatment.

Mammography plays a key role in fighting breast cancer, and research has found that screening has reduced breast cancer mortality by up to 44%. However, low-dose x rays will generally reduce the contrast of the mammograms. In order to resolve this issue, Chapter 4 introduces a new nonlinear unsharp masking (NLUM) scheme for enhancing suspicious regions in mammograms. The NLUM method offers users the flexibility to embed different types of filters in the nonlinear filtering operator, to choose different linear or nonlinear operations for the fusion processes, and to optimize the NLUM parameters manually or by using a quantitative enhancement measure. The chapter also introduces the new second-derivative-like measure of enhancement (SDME). The comparison and evaluation of enhancement performance demonstrates that NLUM can improve the disease diagnosis by enhancing suspicious regions in mammograms with no a priori knowledge of the image contents.

Skin cancer is the most common of all cancers, accounting for nearly half of all cancers in the United States. Automatic detection of skin cancer is a key technology in computer-aided skin cancer diagnosis. Chapter 5 studies skin lesion detection based on color information. Several color spaces are studied, and the detection results are compared. Experimental results show that the YUV color space can achieve the best performance. Furthermore, the chapter develops a distance-histogram-based threshold selection method that is proven to be better than other adaptive threshold selection methods for color detection. Based on the aforementioned methods, a hybrid skin-lesion detection algorithm is presented. The book chapter also investigates GPU techniques for skin lesion extraction, and the results show that GPUs have potential applications in skin lesion extraction.

Chapter 6 presents an incremental learning method for lesion detection using endoscopy videos. With advances in data acquisition technology, data has become large and dynamic. A large number of examples often reduces the
generalization error of the trained model. In the deployment of new image-based diagnosis tools such as capsule endoscopy, new examples continue to be acquired, which enriches the understanding of the imaging modality and could potentially alter previous beliefs. Therefore, efficient and scalable learning approaches are needed that can modify the model structure without having to revisit all of the previously processed examples. The incremental learning method presented in this chapter is developed based on geometric support vector machines (SVMs). The chapter describes the concept of the skin of convex hulls and a method to identify it (only the examples within the skin are retained in the incremental training, which is approximated with the extreme points). The set of extreme points are found via a recursive process by searching along the direction defined by a pair of extreme points. When additional examples become available, they are used along with the retained ones within the skin of the convex hull constructed from the previous data set. This process results in a small number of instances used in incremental training steps and, hence, improved memory efficiency to handle a large amount of data, as well as robustness that exhibits competitive performance.

Chapter 7 provides a comprehensive review of a melanoma screening system, including various imaging technologies, publicly available skin lesion data sets, and image analysis methods such as lesion segmentation, feature extraction and selection, and classification. This chapter also describes in detail a method to bridge the gap between the domain knowledge of physicians (i.e., dermatologists) and computer-generated features representing size, shape, spatial relationship, and texture. Comprehensive comparison using publicly available skin lesion data sets demonstrates the advantage of incorporating domain knowledge.

Microcalcifications are tiny deposits of calcium that appear as small, bright spots on mammograms, and the detection of microcalcifications is an extremely challenging task. In Chapter 8, a novel, hybrid 2D complex-wavelet-transform-based (2D-CWT-based) multifractal feature extraction system is proposed for the detection of microcalcification clusters (MCCs) in digital mammograms. A hybrid feature set, including a set of texture-based features and a set of 2D-CWT-based multifractal features, is presented as the input to a SVM classifier for the detection of the MCCs. The 2D-CWT algorithm and its 2D-CWT-based novel multifractal feature extraction scheme are proposed in the book chapter. Experiments demonstrated a good MCC detection rate and a satisfactory ratio of the true positive fraction to the false positive fraction. The proposed MCC detection system with hybrid features provides an adequate framework for MCC detection.

Chapter 9 focuses on challenges in accurately and automatically detecting and validating suspected prostate cancer lesions in biopsy images. Despite recent improvements in detection and treatment, prostate cancer continues to be the most-common malignancy and the third-leading cause of cancer-related mortality in American men. Evaluation of prostate cancer can be
divided into detection, localization, classification, grading, and staging; accurate assessment is a prerequisite for optimal clinical management and therapy selection. Current diagnosis of prostatic adenocarcinoma is conducted by experienced pathologists using visual analysis of biopsy tissue samples: pathologists assess glass slides under a microscope in order to detect the presence of tumors and to assign a grade according to the architecture of prostatic glands. However, the grading process is time-consuming and error-prone, as well as highly influenced by pathologist experience, pathologist fatigue, and variability in the image interpretation. Therefore, CAD prostate cancer diagnosis has been developed to assist pathologists in the analysis of histopathology images. Prostate biopsy imaging has been accepted as a primary imaging modality for evaluating prostate cancer grades. In the coming decade, the main aim for prostate cancer imaging is more-accurate disease description, characterization, and interpretation through the synthesis of functional, anatomic, and molecular imaging information; therefore, in order to make accurate diagnoses, it is important to thoroughly understand their advantages and limitations, histological background related with image findings, and their clinical relevance for evaluating prostate cancer. Chapter 9 provides an overview of the current clinical approach for detecting and grading prostate cancer and describes the current status and future potential of CAD technology applied to prostate cancer, which is intended to be a support tool in cancer diagnosis and management. The chapter also presents some future perspectives and new strategies in pursuit of better prostate cancer CAD systems, and so on.

Chapter 10 investigates mass analysis using fractal dimension and shape factors in order to differentiate benign masses and malignant tumors. Fractal dimension (FD) and several shape factors—including compactness, convex deficiency, a measure based on Fourier descriptors, fractional concavity, and spiculation index—were calculated from the contour of a mass and used to estimate whether the mass is benign and malignant. The results indicate that shape analysis can lead to efficient discrimination between benign breast masses and malignant tumors. The results also show that fractional concavity gave the highest individual AUC (the area under the receiver operating characteristic curve).

Chapter 11 deals with tomographic imaging, and a new approach for reconstructing images from a finite number of projections is presented. In the new approach, the ray integrals of the image are transformed uniquely into the ray sums of the discrete image on the Cartesian lattice. The experimental results of image reconstruction from a finite number of projections are illustrated in the book chapter and demonstrate the effectiveness of the proposed approach.

Jinshan Tang
Sos Agaian
November 2013
# List of Contributors

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mohamed Abouelenien</td>
<td>University of North Texas, USA</td>
</tr>
<tr>
<td>Sos Agaian</td>
<td>University of Texas–San Antonio, USA</td>
</tr>
<tr>
<td>Kwabena Agyepong</td>
<td>Alcorn State University</td>
</tr>
<tr>
<td>Antonietta Ancona</td>
<td>San Paolo Hospital of Bari, Italy</td>
</tr>
<tr>
<td>Shantanu Banik</td>
<td>University of Calgary, Canada</td>
</tr>
<tr>
<td>C. L. Philip Chen</td>
<td>University of Macau, China</td>
</tr>
<tr>
<td>Loredana Chieppa</td>
<td>University of Bari, Italy</td>
</tr>
<tr>
<td>Balathasan Giritharan</td>
<td>City Grid Media, USA</td>
</tr>
<tr>
<td>Artyom M. Grigoryan</td>
<td>University of Texas–San Antonio, USA</td>
</tr>
<tr>
<td>Yanliang Gu</td>
<td>Michigan Technological University, USA</td>
</tr>
<tr>
<td>Samuela L’Abbate</td>
<td>University of Bari, Italy</td>
</tr>
<tr>
<td>Jianguo Liu</td>
<td>University of North Texas, USA</td>
</tr>
<tr>
<td>Fabio Mangieri</td>
<td>San Paolo Hospital of Bari, Italy</td>
</tr>
<tr>
<td>Clara Mosquera-Lopez</td>
<td>University of Texas–San Antonio, USA</td>
</tr>
<tr>
<td>Issam El Naqa</td>
<td>McGill University, Canada</td>
</tr>
<tr>
<td>Karen Panetta</td>
<td>Tufts University, USA</td>
</tr>
<tr>
<td>Maria Luisa Pepe</td>
<td>University Hospital – Policlinico of Bari, Italy</td>
</tr>
<tr>
<td>Grazia Raguso</td>
<td>University of Bari, Italy</td>
</tr>
<tr>
<td>Rangaraj M. Rangayyan</td>
<td>University of Calgary, Canada</td>
</tr>
<tr>
<td>Ning Situ</td>
<td>Microsoft, USA</td>
</tr>
</tbody>
</table>
Jinshan Tang  
Michigan Technological University, USA

Ian Thompson  
University of Texas Health Science Center, USA

Alejandro Velez-Hoyos  
Pablo Tobon Uribe Hospital, Colombia

Jianhua Yao  
National Institutes of Health, USA

Xiaohui Yuan  
University of North Texas, USA

Xiaojing Yuan  
University of Houston, USA

Ping Zhang  
Alcorn State University, USA

Yicong Zhou  
University of Macau, China

George Zouridakis  
University of Houston, USA