

# Bilateral Comparison of Breast Tissue Density by Using Grey-Level Statistics in 2D Mammograms to Identify Regions of Abnormal Mass (Progress Report)

Jonathan Francis Roscoe  
jjr6@aber.ac.uk  
Computer Science  
University of Wales, Aberystwyth

November 26, 2009

*This report is a summation of my current research and experimentation into approaches for the detection of abnormal masses in digital mammography for the purpose of my third year undergraduate dissertation.*

---

## 1 Project Summary

Breast cancer is the most common form of cancer in the United Kingdom, resulting in nearly 12000 deaths annually in the UK. Breast screening has proven to be an effective means of detecting early stage breast cancer in asymptomatic subjects [10], in several countries including the UK, Finland and Spain [37] [4] national screening programmes are now in full effect. The process of screening involves taking x-rays of each breast (referred to as a mammogram), the resulting image may display present various indicators of abnormal masses in the breast tissue. This enables physicians to discover malignancies in a presently asymptomatic patient. According to publications from Cancer Research UK [39] and the NHS [30], breast screening saves approximately 1400 lives every year but around 20% of screens result in a false-negative (when a patient is mistakenly told they have no dangerous masses, false-positives can occur). Diagnosis of a radiogram subject can be an arduous task and due to the indeterminate nature of mammograms, two diagnosticians will often have different opinions about what they see. Analysis is subject to error, sometimes by human oversight and misjudgement other times by technical inadequacies [34]. Computer Aided Diagnosis has demonstrated the ability to improve detection of malignancies and reduce the occurrence of inaccurate results.

The female breast is constructed of a collection of tissues including adipose, muscle and glandular, and contains a large concentration of blood vessels and lactiferous ducts (see Figure 1). This results in a varied appearance when viewed on a mammogram (see Figure 2) and it can often be difficult to determine if an abnormality is present. In radiology it is widely accepted that the breasts should be treated as a symmetrical organ [24]; although this does not mean that we can expect x-rays of left and right breasts to appear identical (even with perfect positioning of the subject), it does mean that we can expect to see similar density of breast tissue as well as a similar size and structure.

The project is a research undertaking to investigate a new technique of detecting regions of abnormal mass in 2D mammograms using a method bilateral comparison of grey-level intensities. By comparing average deviations in light intensity on mamograms of both the left and right breast of a subject, we hope to be able to detect abnormalities based on pixel regions of extreme light intensity. Due to the nature of x-rays, brightness has a direct correspondence to tissue density, and masses frequently manifest as abnormally dense. This is explained in more detail later in this document.

The material involved is exceptionally interesting due to the complexity and variety of technologies and the scientific understanding required. Although the main focus is in computer vision, knowledge of visualisation, artificial intelligence and radiographic technique are necessary to effectively complete the project. Aside from technical issues there are complex practicality, ethical and affectivity concerns associated with systems of this nature.

### 1.1 Technical Aspects

Over recent decades computer technology has significantly improved the techniques of medical imaging, for example, Computer Tomography (CT) imaging relies on computer processing to visualise two dimensional x-rays,

however, actual analysis of image data has been left to human diagnosticians. Increasingly, technologies enabling computer analysis of these images are emerging. Computer Aided Diagnosis (CAD) is a fast developing and multi-faceted interdisciplinary field, drawing on a number of computer technologies such as artificial intelligence, image processing and visualisation, as well as related radiological, diagnostic and similar medical advancements. All of which makes CAD a very compelling topic and a field which one should be proud to participate in.

## 1.2 Ethical and Affectivity Concerns

Although this is purely a research project with no practical application intended, it is wise to consider the consequences of such a system in the real world.

Computer Aided Diagnosis is an increasingly common technology in medicine and has been used in practical diagnosis for some time [6]. However, a paper by Given-Wilson et al [34] concerning the causes of false-negative

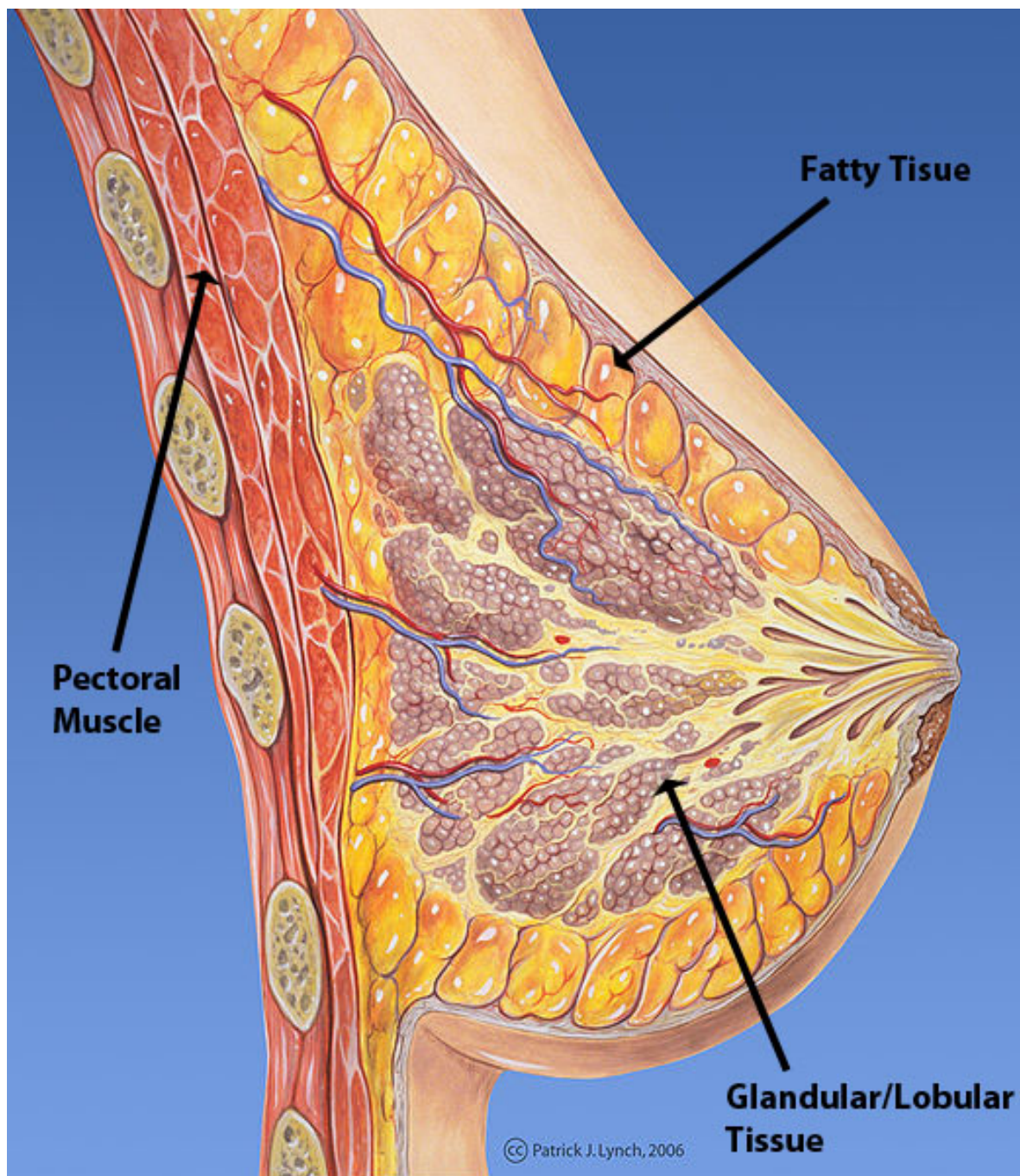


Figure 1: This illustration shows the variety of tissue within the breast from an mediolateral oblique viewpoint. There is a distinct region of glandular tissue, surrounded by fatty tissues. Note the presence of the pectoral muscle, which is typically present in a mammogram. *Source: Patrick J. Lynch, medical illustrator*

diagnosis by human diagnosticians points out that the cause of mistakes is not always a result of oversight, but can be due to technical problems such as malconfigured cameras; in such a scenario CAD systems are unlikely to perform significantly better due to the source being problematic. Consequently CAD systems should not be relied upon absolutely, although we would hope that a modern CAD system would be capable of detecting inadequate data sources. If we accept that a CAD system cannot be infallible then we must consider the repercussions of misdiagnosis.

As the presence of breast cancer is often an exigent and traumatic discovery it is paramount that patients are given the utmost consideration before being presented with a diagnosis – good or bad. A false-negative diagnosis can delay important treatment, whilst a false-positive diagnosis can result in emotional stress and undue medical intervention. A non-conclusive diagnosis can be just as bad, either giving the patient false



Figure 2: This is a typical example mammogram from the MIAS database. Note that as in Figure 1 there is a clear region of glandular tissue, which appears brighter and more textured than the surrounding fatty tissue as a result of varying density. Note also that the pectoral muscle has a similar luminosity to the glandular tissue, this is important to consider because it can make segmentation of the glandular tissue awkward. Also shown is a small plate that has been used to label this particular x-ray, this can cause some difficulty for us as it may be picked up by edge finding features of our software.

security or resulting in further, more invasive examination.

There is a 3 step assessment process commonly used to determine with high certainty the presence of a suspected malignant mass in breast tissue without the need for major surgery [26], this is:

1. Palpate examination by medical professional
2. X-ray and/or ultrasound mammographic analysis (radiomammogram is usually preferred)
3. Needle biopsy (fine-needle aspiration and/or core needle biopsy)

It was the original intention of the project that to ensure the most accurate diagnosis the software should alert the diagnostician and recommend a needle biopsy in the event of any mass detection. This is a naive suggestion however, particularly when no consideration is given to more detailed image assessment first. Therefore, after considering the potential for harm an inadequate diagnosis could cause; it is the opinion of this project that the best action for Computer Aided Diagnosis products is to merely highlight regions of interest and leave treatment or further investigation at the discretion of the human diagnostician. Hopefully, technology will one day be expansive enough for full diagnosis but until then it is best that software be treated as only an aid.

## 2 Background

Computer Vision in medical diagnosis has been a hot topic for some time as there are many visual cues available for diagnosis, this is the basis of CAD. Often CAD is seen as a means to one day eliminate human errors by using machines for a consistent and absolute diagnosis as well as providing an easier and cheaper means of screening. This project does not assume that replacement is currently a feasible option, due to ethical, technical and affectivity concerns, but instead considers computer aided diagnosis as a means of improving human diagnosis, rather than replacing it.

### 2.1 Mass Screening

One of the primary reasons for the development of CAD tools is to enable easier and more efficient mass screening programs. Breast cancer screening is the examination of asymptomatic subjects as a means of detecting early stage breast cancer, the simplest form of screening involves a physical assessment for palpable lumps, however more commonly mammogram screening is the preferred method of assessment due to superior detection rates, increasing availability and costs. Currently only a subset of demographics judged to be 'at risk' (typically post-menopausal women, there are a number of reasons for this including technical issues with tissue density variations over age [29]) are involved in screening programs, however, the availability of accurate and reliable computer aided diagnosis tools could speed up mammogram assessment, decreasing time and monetary costs; enabling programmes to expand their intake and provide a screening programme to a wider population. The most common research seeks to do this by providing a supporting system and enhancing efficiency whilst reducing occurrence of false results.

A significant amount of academic and commercial research and development has already gone into the area of CAD and many techniques have been developed to utilise computer abilities for the purpose of diagnosis. This project is an investigation of the common method of light intensity combined with symmetrical comparison as a means of further improving a computer's ability to locate abnormal masses based on light.

### 2.2 Research in Digital Mammography

The Mammographic Image Analysis Homepage [27] lists approximately 40 different research groups around the world, among them is the Department of Computer Science at Aberystwyth University. The wider field of digital mammography has several major publications and consortiums, most notably the International Workshop on Digital Mammography (IWDM) [18], a meeting held every two years since 1992 for those "who are jointly committed to developing technology, not just for its own sake, but to support clinicians in the early detection and subsequent patient management of breast cancer", the proceedings of which have been most useful for this project.

#### 2.2.1 Notable Publications

A number of useful scientific papers and books relating to digital mammography and vision have been published, covering, comparing and documenting varied techniques for digitising, analysing and interpreting 2D mammograms, as well as more general papers on radiology. The most pertinent publications for formalising, planning and developing this project include:

- Early Stage Breast Cancer: From screening to multidisciplinary management [26]

An informative medical text book that in the early chapters introduces the concepts of screening as well as the appearance of different forms of mass in breast imaging leading onto treatment and management strategies. Whilst information on CAD technology is sparse, it proved a useful aid for understanding the modus of breast cancer.

- Breast Imaging [24]

An easily accessible but thoroughly technical text book that describes methods of breast imaging and the appearance of masses in further, more precise and annotated detail that "Early Stage Breast Cancer" this has been the most useful resource for information on indicators and assessment of malignant masses.

- Learning OpenCV [3]

Whilst many popular computer vision books are theoretical in nature, O'Reilly's "Learning OpenCV" provides a practical introduction to solving vision problems using the OpenCV library. As this project makes use of the OpenCV library the book is wise accompaniment during the development process. It is written by the researchers at Willow Garage [11] and begins with simple examples building up to complex image processing and machine learning.

- Automation in Mammography: Computer vision and human perception [1]

This paper from Departments of Medical Biophysics and Psychology, University of Manchester considers a number of approaches to the implementation of a computerised mass detection system. A framework methodology is described and five key areas are outlined for consideration in the implementation of a system: "identifying the most appropriate targets for assistance, determining performance standards for different tasks, defining a mechanism for demonstrating clinical acceptability, developing new techniques for the automatic detection and analysis of signs of abnormality, and determining the effects of computer-based aids on human performance".

- False negative mammography: causes and consequences [34]

An investigation into the occurrence, causes and consequences of false negative results in mammogram analysis. A series of cases with negative results were given a second review and the results compared, almost two thirds of cases were shown to be false negative results on second diagnosis. The paper considers some of the difficulties in analysing mammograms and demonstrates the need for improvement in the area of diagnosis.

- Mammographic density: a heritable risk factor for breast cancer [8]

There is a strong correlation between mammographic tissue density and an increased risk of breast cancer. This paper investigates the genetic factors of breast tissue density and whilst not directly related to mass detection contains a variety of citations on fringe topics.

- Automatic Mass Segmentation in Mammographic Images [13]

This is the PhD thesis of Oliver Arnau that documents the implementation of a set of tools for detecting early signs of breast cancer in mammographic images. The paper includes a concise review of a variety of techniques for mass detection and outlines a mass detection framework based on learned pattern matching. The methods discussed share much similarity with my own work which makes it an ideal resource.

- Automatic Registration of Mammograms Using Texture-based Anisotropic Features [43]

A discussion on a proposal for an automated registration framework to identify the differences between corresponding mammographic images. The paper investigates detection of abnormalities through the matching of texture features and of particular interest is discussion of methods for segmenting the breast region from an image.

- A Review of Automatic Mass Detection and Segmentation in Mammographic Images [31]

This currently unpublished paper is a review of well known approaches to computer aided detection of masses in mammographic images. The most pertinent section of the paper discusses mass detection using multiple views, comparing temporal, bilateral and ipsilateral methods, of which bilateral is the most popular.

- One Year of Experience with Remote Quality Assurance of Digital Mammography Systems in the Flemish Breast Cancer Screening Program [20]

Also presented in the IWDM 2008 workshop, this paper is not specific to CAD but looks at issues of quality control in for digital mammography systems. These issues include image exposure, artifacts and other image factors. This article is of great interest because of the considerations for accuracy and reliability outlined.

- Comparison of Multiple View Strategies to Reduce False Positives in Breast Imaging [21]

A comparison of multiple viewing strategies for mass detection, the basis of this paper is that current mass detection systems are over sensitive and produce a high number of false positives. The conclusion of the paper is that although multiple view strategies alone are not as effective as single view methods, it is advantageous that they provide complementary information.

- Identification of Regions of Interest in Digital Mammograms [36]

The authors of this paper investigated clustering and gradient based techniques for discovering regions of interest in mammograms. The writing considers the variation in grey-level values across varying tissue and notes the difficulties of segmentation due to subtle variance. The results of the project discussed show that region growing methods can yield a greater number of regions of interest.

A larger list of related materials can be found in the bibliography at the end of this document.

## 2.3 Existing Commercial Developments

As is the case of many research fields, computer vision has stepped out of the laboratory and taken a more practical nature, often becoming commercialised; this is well demonstrated in digital mammography. iCAD [14] is a prime example, vendors of technology for mammogram, breast/prostate MRI and tomosynthesis, they market a product named 'SecondLook Digital' [15], software that can be combined with the hardware of specific manufacturers to "automatically identify and clearly mark suspicious cancers". iCAD promote the following advantages [15]:

- Offers unmatched productivity to reliably handle high case volumes
- Delivers maximum sensitivity and optimal performance to enhance patient care
- Helps detect up to 72% of actionable missed cancers to reduce oversight error
- Integrates seamlessly with existing systems to improve workflow
- Achieves payback in as few as five months with reimbursement support

Similarly, image diagnost international [16] and HOLOGIC [12] (one of the current sponsors of the IWDM [18]) both offer digital mammography solutions that include a CAD feature for the purpose of "second review".

From these examples we can see that CAD is emerging as a mechanism for providing a form of easily accessible second opinion to diagnosticians. As was surmised earlier in this paper, it seems there is not yet a product purporting to be a complete solution to digital mammography, merely aids. Therefore there is still opportunity for mammogram analysis techniques to be expanded, enhanced and explored.

## 3 Masses in mammograms

This section looks at the appearance and classification of typical masses in mammograms as well as the methods used by CAD systems to detect them.

### 3.1 Types of Mass and Their Appearance

There is great variety in the appearance of masses and unfortunately malignant masses are most difficult to discern [24] [25] due to being spiculated, poorly circumscribed and irregular. In comparison, non-malignant masses such as fibroadenomas and benign cycts often have a sharp bordered appearance and circular shape. The variation of masses can be seen in Figure 3. As a general rule, most masses will have a greater light intensity in an x-ray which is why grey-level based analysis is such a common part of mass detection.

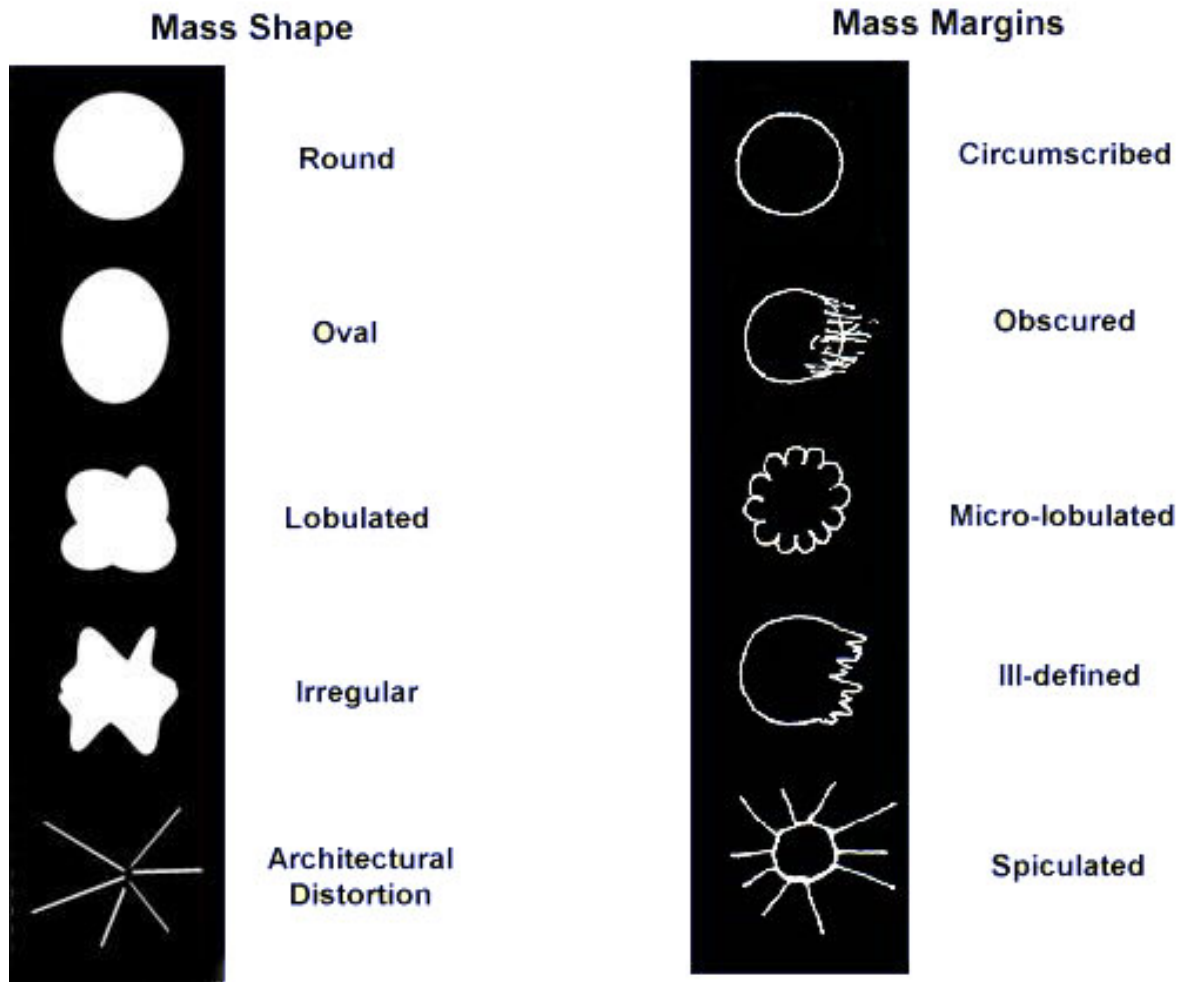


Figure 3: This figure demonstrates the variation in shape and outline of masses as they appear in an x-ray. Source: BB Kopans Breast Imaging [24]

## 3.2 Documented Methodologies for Mass Detection

The basis of detecting masses in mammograms is that pixels forming a mass are non-uniform when compared to a like area of normal tissue. Segmentation of abnormal regions of pixels is often done by detecting variances in pixel intensity (brightness) or texture and may include holistic interpretation of regions.

There are a number of approaches to mass detection in mammograms that have been previously compared [31], each with merits and faults. The following is a brief review of the major areas of research in mass detection:

### 3.2.1 Mass Segmentation

The following techniques are commonly used to detect masses in digital mammograms:

#### Model based

Model based segmentation uses prior knowledge of structures to determine if they are present within the image. Systems implementing this methodology typically include a training stage, where the system is presented with cases where known abnormalities were present, so it can identify homogenous features of masses.

#### Region based

Region based segmentation identifies regions of an image with uniform features, typically, this method is used as a means of extracting masses from the background (normal, uniform tissue).

#### Contour based

Contour based segmentation is used to detect region boundaries within images, ostensibly the aim of such methods is find the edge of masses, though this can be non-trivial due masses being poorly circumscribed.

## Clustering

Clustering based segmentation groups non-adjacent pixels with similar properties to find areas that belong together, once again allowing for elimination of the background.

### 3.2.2 Multiple Views

Multiple views can be used to provide more precise results when combined with data from prior mass segmentation and typically take one or more of the following forms:

#### Temporal

Images are taken from the same physical view, but at different times. This can be used to see if a suspected mass has developed.

#### Ipsilateral

Images of the same breast are taken from both an MLO and CC viewing angle, these images can be combined and lined up for a better three dimensional representation to help overcome issues of noise.

#### Bilateral

Bilateral comparison relies on the basis that the breasts are structurally symmetrical organs and involves comparing images of the left and right breast from the same viewing angle. The most notable implementation is bilateral subtraction [44] which involves the normalisation of two images of each breast from the same viewing angle and then the exclusion of like properties. Research has shown that bilateral subtraction performs better than single image methods [31], though the two are not mutually exclusive.

Multiple views may be used as a means to confirm the presence of masses detected with other mechanisms or may be the means of detection itself (as is the case with bilateral subtraction). The technique proposed by this project uses a method of bilateral comparison, where grey-level values are compared between mammograms to determine abnormal areas of luminosity within the image.

### 3.2.3 Grey-level Intensity

It is important to mention grey-level intensity, also known to as pixel value, brightness or density. This refers to the discrete integer representation of a single pixel and is one of the simplest ways of detecting non-uniform areas. It is an important part of the project methodology.

## 4 Implementation Options and Choices

This section reviews the software platform chosen for this project.

### 4.1 Vision Libraries

A computer vision library is being used for this project due to the complexity of the image processing functions required to implement a CAD system, a library provides a stable and expansive means of quickly developing a new system by drawing on previous efforts. The OpenCV library [33] was chosen due to the available algorithms, structure and general popularity, below is a brief look some of the popular vision libraries:

#### 4.1.1 OpenCV

OpenCV is an open source, cross-platform computer vision library written in C, originally developed by Intel and now supported by Willow Garage [11], it has been used for a variety of applications and has seen contributions from Google and Princeton University, the the library's main focus is on real-time processing. The library is well documented and contains common functions for image analysis, filtering and machine learning, it is constructed of four major components:

- **cxcore** - basic structures and functions
- **cv** - image processing and vision algorithms
- **highgui** - graphical interface, image and video I/O
- **ml** - machine learning library - statistical classifiers and tools

OpenCV 2.0 was released September 2009 which makes it the most recently updated of the libraries.

### 4.1.2 VXL

Written in C++, the Vision "*something*" Libraries [42] are built up of five separate libraries:

- **numerics** - Numerical containers and algorithms. e.g. matrices, vectors, decompositions, optimisers.
- **imaging** - Loading, saving and manipulating images in many common file formats, including very large images.
- **geometry** - Geometry for points, curves and other elementary objects in 1, 2 or 3 dimensions.
- **streaming I/O, basic templates, utilities** - Miscellaneous platform-independent functionality.

The library appears limited in functionality and documentation but development updates have been as recent as September 2009 and new releases are planned for January 2010.

### 4.1.3 Image Processing Library

The Image Processing Library (IPL98) is another multi-platform C/C++ library, the most recent version was released in January 2006. The available vision algorithms are limited to histograms and hough transformations and documentation is unfortunately scarce.

## 4.2 Environment

Linux will be the primary development environment for this project, however, the code is intended to be cross-platform so will not make use of any system specific libraries and every effort will be made to test code in a Microsoft Windows environment when possible.

The software used during testing, development and deployment will be:

#### Linux:

- Ubuntu Karmic 9.10 and CentOS 5.3
- vim
- g++ with make and gdb
- Sun Netbeans 6.7

#### Windows:

- Microsoft Windows 7
- Microsoft Visual Studio 2008
- Scite 2.01
- CMake [32]

The most recent stable release of OpenCV will be used at all times, unless compatibility issues arise.

One of the secondary goals of the project is to develop a web-based front end for the final product, if this is developed the following environment will be used:

- CentOS 5.3
- Apache 2
- PHP 5.3
- MySQL 5

## 4.3 Source Code and Revision Control

All code is being developed using C++98. In the interest of data security, change tracking and code integrity [38], all code and documentation is stored in a private Subversion 1.6 repository backed up to an Aberystwyth University owned filestore. As I am the sole developer for the project I do not require the facilities of an open, distributed revision control system such as git, and chose Subversion specifically for the server based methodology that allows me to easily keep my most recent code in a single master repository.

## 5 Goals and Objectives

The project will be evaluated based on the completion of software capable of analysing mammograms and detecting abnormal masses, in such a way that data on detected masses can be collected and used to determine the effectiveness of the implemented method as discussed in section 8.

The data will then be compared with known masses as specified in the MIAS metadata files. We would hope to see a strong affirmation of mass presence. If the software proves an effective means of mass detection we would offer a web based service for mamogram analysis - diagnosticians could upload image pairs to a website and receive a list of suspicious areas that were detected.

## 6 Process Model

### 6.1 Prototyping

As this is a research project rather than a commercial development the initial development will abandon traditional methodologies, instead there will be the regular development of distinct, standalone prototypes to solve specific problems with no formal development methodology. These prototypes will be small, creative solutions unhindered by formal process and will later be discarded. This is beneficial to me as it allows to experiment in how I solve each individual problem, without needing to worry about integration with final product or extensive refactoring. This is of particular advantage because I am previously unfamiliar with computer vision, the OpenCV library and C++ paradigms so I will not be hindered by mistakes made during the initial stages of learning and familiarisation. This method of takes the following pattern:

1. Planning - break final objective into small tasks and prioritise
2. Iterative Phase part 1 - develop prototype solution for current problem of highest priority
3. Iterative Phase part 2 - review solution, choose next task and return to phase part 1 until all tasks complete
4. Review Requirements - ensure all requirements for overall objective (final product) are met individually by the prototypes created
5. Design and create final product from scratch

This is a list of the prototypes to be developed:

1. tool to open and produce basic data on MIAS images, in preparation for manipulation by OpenCV
2. tool to analyse images and equate average light intensities for a  $x^2$  region of pixels
3. tool to identify fatty/dense/background regions of a mammogram
4. tool to locate like regions on left and right mammograms
5. tool to identify  $x^2$  areas of pixels with an average light intensity beyond the standard deviation for that region

### 6.2 Deliverable/Final Product

In the final stages of the project, knowledge gained in the development of prototypes will be combined to design and develop a cumulative tool that will be able to perform the mammographic analysis as outlined in the project description and generate a list of suspected masses. For this a more formal methodology will be adopted, making use of agile development techniques, in particular, feature driven development.

### 6.3 Testing

Although test driven development is not being employed, for the sake of maintainable and reliable code, the project will endeavour that the final product has 100% code coverage with function based testing.

## 7 Technical Challenges and Considerations

### 7.1 Data Source

When creating mammograms there are four possible viewing angles, left or right of either mediolateral oblique (MLO) or cranial-caudal (CC). For this project we will be using only MLO views as this is the most commonly used. The data being used for the project is the MIAS [5] database of digital mammograms. The images are single channel (greyscale) and each pixel is represented by an unsigned 8-bit value which is a direct representation of light intensity.

The images in this database were scanned with a Joyce-Loebl microdensitometer SCANDIG-3, which has a linear response in the optical density range 0-3.2. Each pixel is 8-bits deep and at a resolution of 50um x 50um. Further details of performance can be found in:

D H Davies (1993) 'Digital mammography - the comparative evaluation of film digitizers' British Journal of Radiology, Vol.66 pp930-933.

### 7.2 Technical Challenges

The primary technical challenges this project presents are the classification of regions within the images and the subsequent statistical comparison of grey level data for each region to the opposite side. A more detailed investigation of the process is described in the "Project Methodology" section.

#### 7.2.1 Image Segmentation

Image segmentation is the process by which distinct regions of pixels are identified from the larger image, typically, this is based upon edge detection algorithms that find regions of an image where there is distinct variation between pixels. Segmentation is most commonly used to find objects in a 3D scene for the purpose of machine vision, however, for this project the segmentation has two objective:

1. create distinct regions for use as 'working areas' which can be further subdivided to calculate average pixel intensity for variable squared areas
2. provide a means of locating corresponding regions on left and right mammograms for the purpose of bilateral comparison

The goal of the segmentation is to detect and classify three distinct regions of the mammogram image: background, lobular tissue and fatty tissue. To the human interpretation, this is presented as three clear regions expanding outward from the centre of the breast, however as machines lack contextual information that would enable them to recognise the regions based on overall structure it may be less clear to the software. The image will then be divided into quadrants, which in an ideal example would result in twelve distinct segments. Quadrants will be numbered in a clockwise fashion starting with the top left for left view MLO, and counter clockwise from the top right for a right view. A perfect example is demonstrated in Figure 7.2.1.

#### 7.2.2 Statistical Comparison

In this project, the detection of abnormal regions is based on a change in grey-level values beyond the standard deviation. The key to performing the bilateral comparison is to calculate average light intensities for segments on left and right images and then search for sub-segments with an average intensity outside of the the standard deviation.

## 8 Project Methodology for Mass Detection

Due to the nature of technology, a conventional 2D x-ray image should be understood to be a dimensionally reduced 3D image. That is to say, although there is no palpable depth to such an image, depth is expressed as a function of light intensity in the image. The opacity of an x-ray image is determined by the ease with which x-rays are able to pass through and denser tissue results in less radiation passing through the subject, which in turn results in brighter regions on the image. Since malignant masses are most often a growth of cancerous cells, they are denser when compared to normal tissue; consequently these masses typically manifest in the x-ray image as bright regions.

Many malignant masses have irregular structures with poor definition [25] which can make them more difficult to locate, due to normal variations in the surrounding tissue. By comparing the densities of like regions on each breast we could determine when one has a particularly abnormal density, enabling us to locate masses

with poorly formed boundaries. This technique could also be combined with other techniques to aid conclusive detection of abnormalities.

The steps for isolating points of abnormal average light intensity are as follows:

1. Select a segment (from the twelve described in the previous section)
2. For the selected segment, calculate standard deviation from the average light intensity
3. Calculate the standard deviation from the average light intensity for the same segment on the opposite image
4. Calculate the average of the two standard deviations
5. For both segments, sequentially search the image for sub-segments of size  $x$  with an average intensity beyond the standard deviation ( $x$  is a value to be determined by experimentation)
6. Compare results, if similar sub-segments are detected for both sides, ignore them
7. Report any detected sub-segments on both images

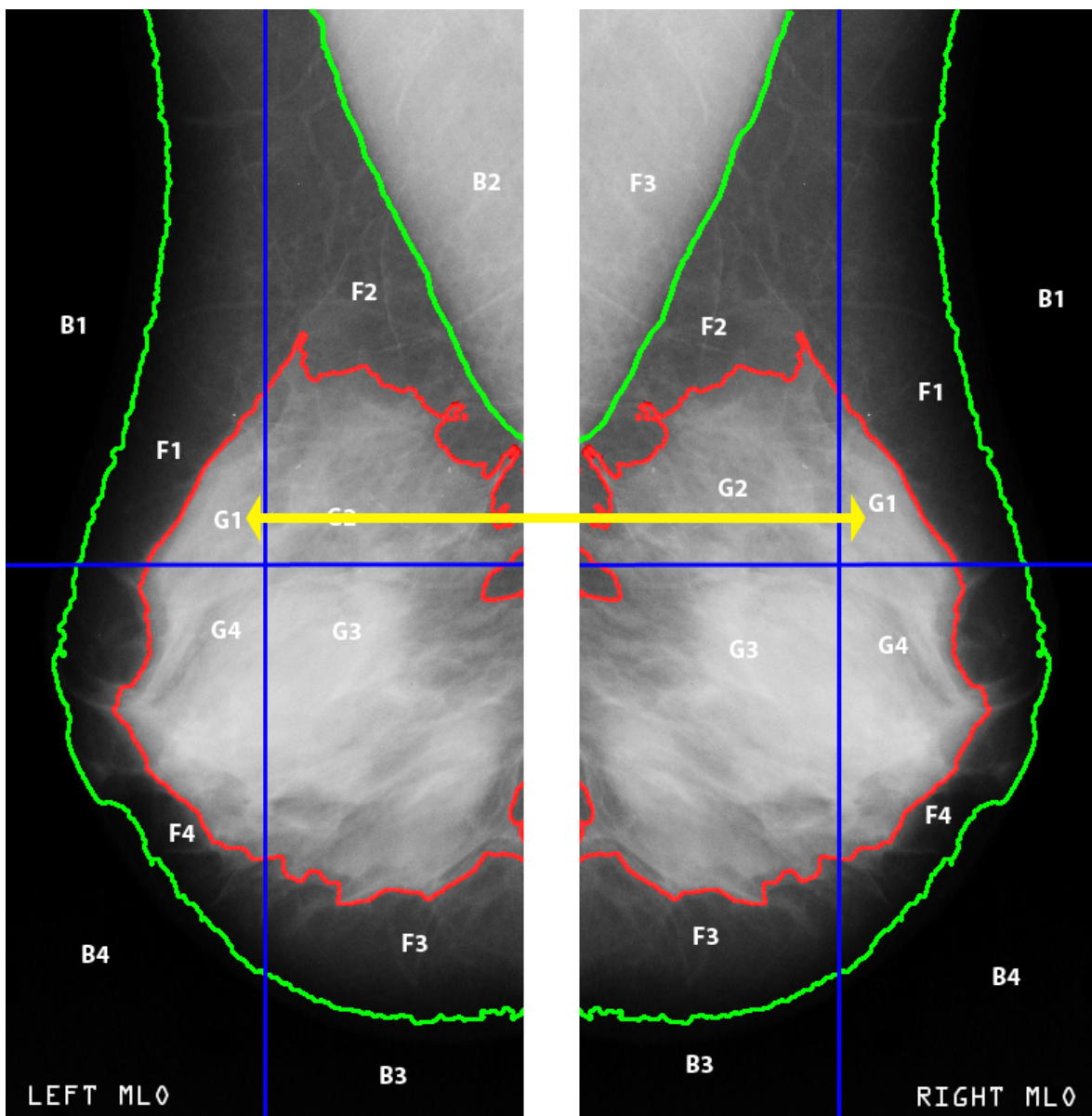


Figure 4: Example of ideal segmentation of mammogram images. Green lines separate the background and fatty tissue, red lines separate glandular tissue and fatty tissue. The blue lines indicate where the image is broken into quadrants. The yellow arrow demonstrates how segments from an MLO view can be compared with its opposite. In reality, samples are less likely to appear as symmetrical.

8. Repeat for all remaining segments

Figure ?? displays the method of segmentation.

## 9 Design Outline

Whilst the final product design is subject to progressions in knowledge and understanding of the opencv library and computer vision design techniques, the current broad design outline for the system has the following five constituents:

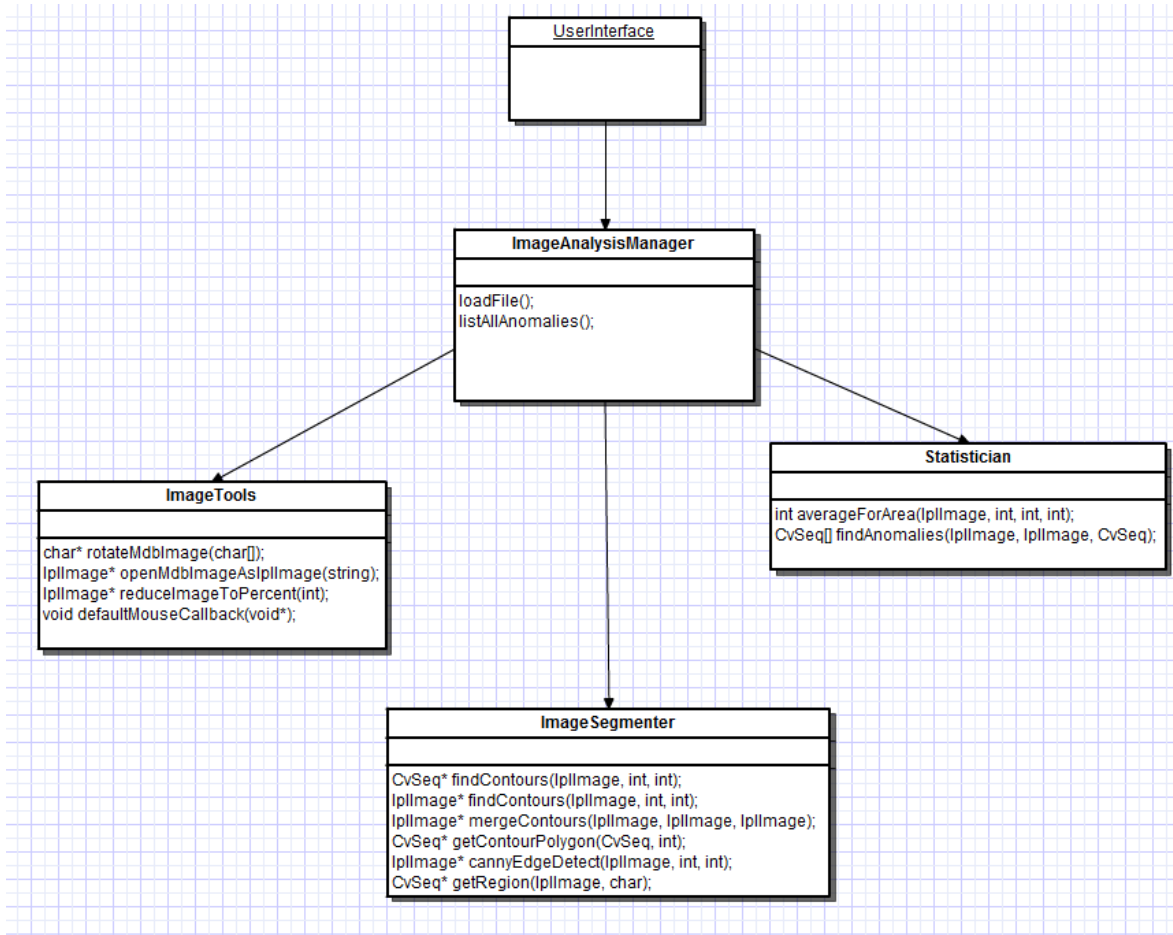


Figure 5: This class diagram outlines the proposed structure of the final product. The UI module is intended to be easily interchangeable and the the ImageManager class is to facilitate communication with fine grained functionality.

- **UI** - user interface, used to take user input and display processed images
- **ImageManager** - provides high level functions for detecting masses
- **ImageTools** - a collection of useful functions for working with the mdb images
- **ImageSegmenter** - contains functions for classifying regions of mammograms
- **Statistician** - contains functions for gathering grey-level statistics on regions

These are outlined in Figure 9.

The MIAS data contains various metadata describing masses for each image. In order to store this important metadata two additional classes will be created - MdbImage and MdbAbnormality. The only difference between an MdbImage and a traditional OpenCV IplImage structure is that it will contain an array of abnormalities which will be populated with MdbAbnormalities as defined in the metadata. The MdbAbnormality class contains properties to hold the location, size and type of abnormality as seen in Figure 9. This data is important in order to create a tool that will compare the results of our program with the known results.

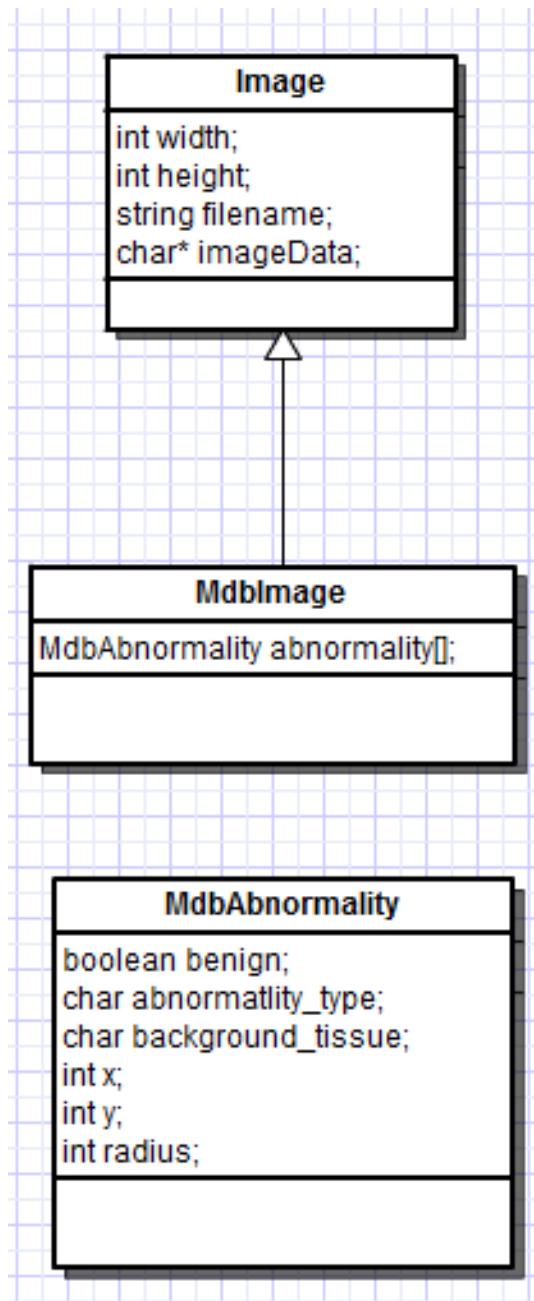


Figure 6: The MdbAbnormality contains information on the size, location and variety of an abnormality within an image. It will be used to compare against the regions marked as suspicious by our software.

## 10 Weekly Plan

Approximately 12 hours per week will contribute to project development. This is a list of the scheduled deliverables that must be adhered to in order to ensure smooth running and success of the project.

Date	Deliverable
06/11/09	Tool to read MIAS mdb image files
13/11/09	Tool to analyse images and equate average light intensities for $x^2$ region of pixels
27/11/09	Tool to identify fatty/dense/background regions of a mammogram
18/12/09	Tool to locate like regions on left and right mammograms
25/12/09	Tool to read MIAS co-ordinate data
05/02/10	Delivery of complete mammographic analysis tool (final product)
26/02/10	Completion of data collection/analysis/testing
26/03/10	Delivery of first dissertation draft and associated presentation material
16/04/10	Completion of presentation material and any other preparations
22/04/10	Dissertation delivery, project end.

## References

- [1] SM Astley, I Hutt, and et al S Adamson. Automation in mammography: computer vision and human perception. Proceedings of SPIE.
- [2] R E Bird, T W Wallace, and B C Yankaskas. Analysis of cancers missed at screening mammography. September 1992 Radiology, 184, 613-617.
- [3] Gary Bradski and Adrian Kaehler. Learning opencv. O'Reilly Media, ISBN 978-0-596-51613-0, 2008.
- [4] Adrain Burton. Europe continues breast screening despite doubts. The Lancet: Oncology by Elsevier Ltd, Volume 3 Issue 5, Page 258, 2002.
- [5] University of Manchester Dept. of Medical Biophysics. Mammographic image analysis society.
- [6] Kunio Doi, Heber MacMahon, Shigehiko Katsuragawa, Robert M. Nishikawa, and Yulei Jiang. Computer aided diagnosis in radiology: potentials and pitfalls. <http://asia.edu.tw/~leecheng/Research/new/97-doi.pdf>.
- [7] 9th International Workshop IWDM 2008 Elizabeth A Krupinski. Digital mammography, 2008.
- [8] Boyd et al Methods in molecular biology. Mammographic density: a heritable risk factor for breast cancer. <http://www.ncbi.nlm.nih.gov/pubmed/19107441>, 2009.
- [9] Boyd et al New England Journal of Medicine. Mammographic density and the risk and detection of breast cancer. <http://content.nejm.org/cgi/content/abstract/356/3/227>, 2007.
- [10] A. P. M. Forrest and R. J. Aitken. Mammography screening for breast cancer. Annual Reviews of Medicine, Vol. 41, pp. 117-132, 1990.
- [11] Willow Garage. Developers of hardware and open source software for personal robotic applications. <http://www.willowgarage.com/pages/software/opencv>.
- [12] Hologic. Hologic breast imaging. <http://www.hologic.com/breast-imaging/diagnostic-workstations/>.
- [13] Oliver i Malagelada. Automatic mass segmentation in mammographic images, 2007.
- [14] iCAD. Industry-leading provider of medical image analysis. <http://www.icadmed.com/>.
- [15] iCAD. Secondlook digital. <http://www.icadmed.com/products/mammography/SecondLookDigital.htm>.
- [16] image diagnost international. image diagnost international. <http://www.imagediagnost.de/>.
- [17] National Cancer Institute. International cancer screening network. <http://appliedresearch.cancer.gov/icsn/>.
- [18] IWDM. International workshop of digital mammography. <http://www.iwdm2008.org/>.
- [19] De Koning H J, Fracheboud J, Boer R AD Verbeek A L, Collette H J, Hendriks J H, van Ineveld B M, de Bruyn A E, and van der Maas P J. Nation-wide breast cancer screening in the netherlands: support for breast cancer mortality reduction. International journal of cancer. 1995; 60(6): 777-80.
- [20] Jurgen Jacobs, Kim Lemmens, Joris Nens, Koen Michelson, Guy Marchal, and Hilde Bosmans. One year of experience with remote quality assurance of digital mammography systems in the flemish breast cancer screening program. Digital Mammography, 9th International Workshop IWDM 2008.
- [21] Melicor Sentis Joerg Fernandez. Comparison of multiple view strategies to reduce false positives in breast imaging. Digital Mammography, 9th International Workshop IWDM 2008.
- [22] S Astley K W Bowyer. State of the art in digital mammographic image analysis, 1994.
- [23] Nico Karssemeijer, Andrea Hupse, Maurice Samulski, Michiel Kallenberg, Carla Boetes, and Gerard dan Heeten. An interactive computer aided decision support system for detection of masses in mammograms. Digital Mammography, 9th International Workshop IWDM 2008.
- [24] Daniel B. Kopans. Breast imaging 3rd edition. [http://books.google.co.uk/books?id=LvU\\_v\\_dP08cC](http://books.google.co.uk/books?id=LvU_v_dP08cC), 2007.
- [25] Valerie P. Jackson Lawrence Wayne Bassett. Diagnosis of diseases of the breast 2nd edition. <http://books.google.co.uk/books?id=keTzVe31kMUC>, 1997.
- [26] G. Querci della Rovere M. W. E. Morgan, Ruth Warren. Early breast cancer: From screening to multidisciplinary management. 2nd edition, 2006.
- [27] mammoimage.org. The mammographic image analysis homepage. <http://www.mammoimage.org/>.

- [28] mcgill.ca. Interactive mammograph analysis web tutorial. <http://sprojects.mmi.mcgill.ca/Mammography/cases.htm>.
- [29] NHS. Nhs breast screening programme faq. <http://www.cancerscreening.nhs.uk/breastscreen/#under50>.
- [30] NHS. Breast screening - the facts. <http://www.cancerscreening.nhs.uk/breastscreen/publications/ia-02.html>, 2009.
- [31] Arnau Oliver, Jordi Freizenet, Joan Marti, Josep Pont, Elsa Prez, Erika R.E. Denton, and Reyer Zwigelaar. A review of automatic mass detection and segmentation in mammographic images. TBA.
- [32] OpenCV. 2.0 readme. <http://sourceforge.net/projects/opencvlibrary/files/opencv-win/2.0/0penCV-2.0.0a-win32.Readme.Please.txt/download>.
- [33] OpenCV. Opencv. <http://opencv.willowgarage.com/wiki/>.
- [34] M. Warren R. Given-Wilson, G. Layer and London J.-C. Gazettn St George's Hospital. False negative mammography: causes and consequences. *The Breast* (1997) 6, 361-366, 1997.
- [35] Marc J. Rochkind. The source code control system. *IEEE Transactions on Software Engineering*, Vol SE-1, No 4, December 1975 - <http://basepath.com/aup/talks/SCCS-Slideshow.pdf>, 1975.
- [36] Reem Al-Mansoori Sameer Singh. Identification of regions of interest in digital mammograms. Department of Computer Science, University of Exeter <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.40.5954&rep=rep1&type=pdf>, 2000.
- [37] Sam Shapiro, Elizabeth Ann Coleman, Mireille Breeders, Mary Codd, Harry de Koning, Jacques Fracheboud/ Sue Moss, Eugenio Pad, Sylvie Stachenko, and Rachel Ballard-Baibash. Breast cancer screening programmes in 22 countries: current policies, administration and guidelines. <http://ije.oxfordjournals.org/cgi/reprint/27/5/735.pdf>, 1998.
- [38] Walter F. Tichy. Design, implementation and evaluation of a revision control system. Department of Computer Sciences, Purdue University, West Lafayette, Indiana 47907, 1982.
- [39] Cancer Research UK. Summary of breast cancer activity in the uk. <http://info.cancerresearchuk.org/cancerstats/types/breast/screening/>.
- [40] Vachon, C. M., Pankratz, V. S., Scott, C. G., Maloney, S. D., Ghosh, K., Brandt, K. R., Milanese, T., Carston, M. J., Sellers, and T. A. Longitudinal trends in mammographic percent density and breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.* 16: 921-928, 2007.
- [41] C. M. Vachon, T. A. Sellers, Pankratz, V. S., F. M. Hall, N. F. Boyd, Hopper J. L., and E Thurfjell. Mammographic density of the breast. *New England Journal of Medicine* 348: 174-175, 2003.
- [42] VXL. C++ libraries for computer vision. <http://vxl.sourceforge.net/>.
- [43] Kexiang Wang, Hong Qin, Paul R Fisher, and Wei Zhao. Automatic registration of mammograms using texture-based anisotropic features.
- [44] Fang-Fang Yin, Maryellen L. Giger, Kunio Doi, Charles E. Metz, Carl J. Vyborny, and Robert A. Schmidt. Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images. *Med.Phys.* Volume18, Issue5, pp. 955-963 (September1991), 1991.