CANCER INTRODUCTION TO THE
DIGITAL HUMAN PHANTOM AND
ITS VISUALIZATION

A progress report submitted to the University of Manchester
for the degree of Master in Advanced Computer Science
in the Faculty of Computer Science

2014

By
Maria Charalampous
School of Computer Science
# Contents

Abstract ................................. 6

Declaration ............................. 7

Copyright ............................... 8

1 Introduction ........................... 9
   1.1 Aims and Objectives .................. 10
   1.2 Deliverables .......................... 11
   1.3 Report Outline ....................... 12

2 Background Study - Lung Cancer 13
   2.1 The Respiratory System ............... 13
   2.2 Lung Cancer General Information .... 14
   2.3 Non-Small Cell Lung Cancer .......... 16
      2.3.1 Squamous-Cell .................... 16
      2.3.2 Adenocarcinoma ................... 17
      2.3.3 Large-Cell-Carcinoma .......... 18
   2.4 Small Cell Lung Cancer ............... 19
   2.5 Lung Carcinoid Tumour ............... 20
   2.6 Detection of Lung Cancer and Current Problems ........... 21
      2.6.1 Chest X-ray ....................... 21
      2.6.2 CT Scans ......................... 21
      2.6.3 PET scan ......................... 23
      2.6.4 Sputum Cytology ................. 23
      2.6.5 Bronchoscopy ..................... 25
      2.6.6 Fine needle aspiration biopsy .... 25
3 Background Study - Digital Human Phantom
   3.1 DHP Production ............................................. 27
   3.2 DHP Images Structure .................................... 29

4 Research Methodology and Project Planning 30
   4.1 Research Methodology ...................................... 30
   4.2 Evaluation Plan ............................................ 31
   4.3 Project Planning ............................................ 32

5 Project Progress 34
   5.1 Meetings with supervisor .................................. 34
   5.2 Understanding of pgm and ppm images ...................... 34
   5.3 Background study on the position of lung cancer ............ 35
   5.4 Shell script for ppm image production ...................... 36
   5.5 Shell script to identify the location of lung in the DHP .... 36

6 Conclusion 37

Bibliography 38

A Shell Script for ppm and pgm image production 42

B Shell Script to identify the location of lung in the DHP 45

Word Count: 6123
List of Tables

3.1 Image numbers ............................................ 29
5.1 PGM file example ........................................ 35
5.2 PPM file example ........................................ 35
List of Figures

2.1 Lungs Anatomy .................................................. 14
2.2 Lung Cancer .................................................... 15
2.3 Squamous Cell Lung Cancer ................................. 17
2.4 Adenocarcinoma Lung Cancer ............................... 18
2.5 Large Cell Carcinoma ......................................... 19
2.6 Small Cell Lung Cancer ....................................... 20
2.7 Chest X-Ray with Adenocarcinoma of the lung .......... 22
2.8 Chest CT Scan with Adenocarcinoma of the Lung ....... 22
2.9 Chest CT scan and PET/CT combination image .......... 23
2.10 Normal Pattern Sputum ....................................... 24
2.11 Adenocarcinoma Lung Cancer Sputum .................... 24
2.12 Squamous Cell Carcinoma Sputum ......................... 24
2.13 Bronchoscopy example ....................................... 25
2.14 CT Scan guided Lung Biopsy ............................... 26

3.1 Human Body MRI scans ...................................... 28
3.2 MRI scan image (v1_00396.jpg) ............................. 28
3.3 DHP image sample ............................................. 29
Abstract

Lung cancer is indicated as one of the most dangerous and common types of cancer which affects people, especially the elderly. It is the leading cause of death among the various types of cancer with its death rate being much higher than any other type of cancer. The current detection mechanisms encounter various problems and there is a need for a better detection mechanism, which will be able to visualize the exact place of the affected lung area clearly based on the healthy human phantom. In order to test novel detection algorithms, numerical simulations need to be performed with the digital human body including lung cancer. This project aims to produce a tool that will replace the healthy lung tissues in digital human phantom with the cancerous cells to represent the areas affected by the lung cancer.
Declaration

No portion of the work referred to in this dissertation has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright

i. The author of this progress report (including any appendices and/or schedules to this progress report) owns certain copyright or related rights in it (the “Copyright”) and she has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this progress report, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the progress report, for example graphs and tables (“Reproductions”), which may be described in this progress report, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this progress report, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations).
Chapter 1

Introduction

According to recent studies, lung cancer [1] is characterised as one of the most dangerous and most common type of cancer that affects people. High rates of lung cancer incidents were detected in older people and especially the elderly. Lung cancer is indicated as the most deadly type of cancer with its death rate being much higher than any other type of cancer such as breast cancer. The three major groups of lung cancer are:

- **Small Cell Lung Cancer** [2]: They are about the 10% to 15% of the cases. Its name was taken because of the small size of the cancerous cell under the microscope. The most common reason that leads to this type of lung cancer is smoking.

- **Non-Small Cell Lung Cancer** [3]: It is the most common type of lung cancer and it occurs in about 85% to 90% of the lung cancer cases. It is very possible to spread in other parts of the body. Non-small cell lung cancer is divided into three categories:
  - **Squamous-cell** [4]: It is the most common lung cancer and it is usually found somewhere in the middle of the lung. Squamous-cell cancer is usually caused by smoking.
  - **Adenocarcinoma** [4]: It appears in the outer parts of the lung. The number of patients with adenocarcinoma increases in such a way that in the near future it might become more common than the squamous-cell lung cancer.
  - **Large-cell-carcinoma** [4]: This type of lung cancer is called large-cell-carcinoma because under the microscope, the cells look very big and rounded. Usually, this type of lung cancer grows very fast.
• **Lung Carcinoid Tumour [2]:** A very rare lung cancer which appears in less than 5% of the lung cancer cases. Another name for lung carcinoid tumour is lung neuroendocrine tumour. Tumours of this category grow slower than tumours in the other categories.

The type of lung cancer is very important because depending on the type there is a different cancer prognosis and different treatment options. The detection of lung cancer is difficult and especially at early stages such as in cancerous cell division. There are several detection mechanisms for lung cancer which vary depending on the type of lung cancer. Such mechanisms are [1]:

- X-ray
- CT Scans
- Sputum Cytology
- Fine needle aspiration biopsy
- Bronchoscopy
- Positron Emission Tomography scan

The current detection mechanisms have various problems and our group is currently developing a new technique to detect the lung cancer at the early stages. Since this is all under research, it is not possible to try our algorithms on real human bodies. Therefore, in order to test the algorithm, a digital human phantom, which is explained in Chapter 3, is needed for the numerical simulations. Unfortunately, all the available digital human phantoms are produced from healthy humans. Therefore, modifications have to be made to the currently available healthy digital human phantom for our own purpose.

### 1.1 Aims and Objectives

This project aims to produce a tool that will replace the healthy lung tissues in digital human phantom with the cancerous cells to represent the areas affected by the lung cancer. In order to get away from the current problems of detection, a better detection mechanism is needed, which will be able to visualize the exact place of the lung cancer clearly based on the healthy human phantom. In order to test novel detection
algorithms, numerical simulations need to be performed with the digital human body with lung cancer in it.

To manage to meet the project aims these have to be done:

- Understand the portable grey format of images
- Learn shell scripting in Linux Operating Systems
- Survey the various types of lung cancer, their location and their characteristics
- Develop an algorithm to map the distribution of the cancer cells to the healthy digital human phantom and to replace the healthy tissue with the cancerous cell
- Implement a visualization tool to show the lung cancer position
- Evaluate the performance of the visualization tool
- Optimize the shell script to have a more efficient computation

1.2 Deliverables

The final artefact that will be delivered in September will include:

**MSc thesis**: The final report which will include all background work done, the methodology followed to create the visualization tool and the evaluation of the tool. Firstly, the introduction will present the purpose of the dissertation and will state the aim and objectives of this study. After that, there will be a background study with a detailed survey on the types of lung cancer and their detection mechanisms. Furthermore, the methodology to build the visualization tool will be described. Next chapter will be the evaluation of the tool and its results. Finally, any conclusions or further work will be stated.

**Shell script and README**: The shell script will basically be the visualization tool and a README file will explain the usage of that script.

**Digital human phantom with lung cancer in it**: The healthy digital human phantom will be modified in order to create the cancerous digital human phantom. Such phantoms will be the outcome of our shell script.
1.3 Report Outline

This progress report includes the following sections:

Chapter 2 - Background Study - Lung Cancer: This chapter presents a detailed explanation of the basic types of lung cancer.

Chapter 3 - Digital Human Phantom: Digital Human Phantom is explained in Chapter 3.

Chapter 4 - Research Methods and Project Plan: In this chapter there is an explanation of the methodology followed in this project, the evaluation plan and the project plan which is adopted for the fulfilment of this dissertation.

Chapter 5 - Project Progress: Chapter 5 states the progress made so far according to the deliverables and the project plan.

Chapter 6 - Conclusion: This chapter summarizes the main points of the project.
Chapter 2

Background Study - Lung Cancer

Cancer [5] is, unfortunately, a well-known term for a disease that counts thousands of peoples deaths every year. The characteristic of this disease is the fast growing rate of cancerous cells and the spread of abnormal cells that cannot be controlled. This project focuses on lung cancer, a specific type of cancer that affects the lungs of a human as its name states.

2.1 The Respiratory System

In order to understand where the cancer evolves in our lungs it is good to have an understanding about the structure of the lungs and how they work. Lungs are two sponge-like organs which are filled with air. They are located in a human’s chest as shown in Figure 2.1. Each lung is divided into parts called lobes. The left lung is divided into two lobes whereas the right lung is divided into three lobes. The left lung is smaller because of the space that is taken by the heart, which is more in that side of the body. [6]

When a human inhales air, it passes through the trachea into the lungs and then the air spreads with the help of the bronchi. The bronchi are divided into smaller branches called the bronchioles which end in tiny air sacs, the alveoli, where the air finally end up. In the alveoli, blood absorbs the oxygen contained in the air and gives the carbon dioxide to the alveoli to be exhaled. Finally, there is a thin tissue layer that covers the lungs, which is called the pleura. The pleura protect the lungs and help them move. [7]
2.2 Lung Cancer General Information

Lung cancer [8] is one of the most common and dangerous types of cancers for both men and women. The number of incidents is very high each year and puts lung cancer in the second place after breast cancer and prostate cancer for women and men respectively. However, lung cancer has the highest death rate among all types of cancer which is way higher than the combination of breast, prostate and colon cancer.

Lung cancer is located in one or both lungs when abnormal, uncontrolled cancerous cells start to grow and create tumours which obstruct the lungs from functioning properly [9]. Most of the times, those tumours start to grow in the walls of the bronchi as seen in Figure 2.2. There have been identified more than 20 types of tumours which originate in the lung [10]. There are three major types of lung cancer which are small-cell lung cancer, non-small cell lung cancer and lung carcinoid tumour which will be discussed later.

Nine out of ten cases of lung cancer are caused by smoking [11]. Not only active smokers may end up with lung cancer but also passive smokers, who inhale the tobacco smoke from other smokers. Except from smoking, there are other risk factors which increase the risk of lung cancer but in a much smaller amount than smoking. According to Cancer Research UK [11], such factors are:

- Exposure to radon gas
- Exposure to certain chemicals
- Air pollution
- Previous lung disease
- A family history of lung cancer
- Previous smoking related cancers
- Immunodeficiency (weak immune system)

Usually, lung cancers do not reveal themselves through symptoms, especially in the early stages of lung cancer. There are though some symptoms that should alarm people and need to be reported to a doctor. Such symptoms, according to the American Cancer Society [12], include:

- A persistent cough which may get worse
- Chest pain
- Hoarseness
- Weight loss
- Coughing with blood or phlegms
• Shortness of breath
• Tiredness
• Persistent infections such as pneumonia or bronchitis

2.3 Non-Small Cell Lung Cancer

The most common type of lung cancer is Non-Small Cell Lung Cancer (NSCLC) [13] with a percentage of 85% - 90% of lung cancer incidents. Although it begins in the lungs, it is very possible to metastasize to other body organs. NSCLC is divided into 3 main subtypes which are called Squamous-Cell Lung Cancer, Adenocarcinoma and Large-Cell-Carcinoma. Despite the fact that they are grouped under the same type, their cells differ in shape, size and chemical make-up. However, the prognosis and treatment are most of the times alike.

The initial step of NSCLC is when the epithelial cells start to grow really fast and cannot be controlled. Many times these cells become masses, known as tumours. Those tumours may stay where they are or spread to other parts of the human body. If a patient is diagnosed with NSCLC in an early stage with the cancer not metastasized to other parts of his body, then the cancer is more easily treatable. Some treatments, that may be applied on the patient to cure NSCLC, are a combination of chemotherapy, radiation and surgeries. [3]

2.3.1 Squamous-Cell

Squamous cell lung carcinoma is a subtype of NSCLC, making up about 30% of NSCLCs, according to Lynne Eldridge [14]. This type of lung cancer is created by the round cells that take the place of damaged cells in the bronchi. Tumours of this type usually appear in the middle of the lungs or in the main airway branches as shown in Figure 2.3. Squamous cell carcinoma can metastasize to other parts of the body such as the bones, the liver or even the brain. Squamous cell lung cancer takes a long time to grow and may need many years to develop to an invasive cancer. Symptoms are very similar to the ordinary cold and they develop very slowly. This type of cancer can appear to be pneumonia or a collapsed lung. [15]

The main cause of squamous cell lung cancer is smoking. Some secondary factors include age, family history, and exposure to smoke or dust. The period a person has
been smoking and the amount of cigarettes he smokes play a very important role in the chances to develop squamous cell lung cancer. If he stops smoking then the chances to develop the squamous cell carcinoma start to decrease. [15]

Squamous cell lung cancer incidents started decreasing the last years. As mentioned by Lynne Eldridge [14], the cause of this downward trend is the introduction of filtered cigarettes to the market. These filtered cigarettes allow the smoke to pass through the bronchial tubes and get deeper into the lungs, where other lung cancers occur, like adenocarcinoma.

### 2.3.2 Adenocarcinoma

According to the College of American Pathologists, adenocarcinoma [16] is the most common type of lung cancer both for smokers and non-smokers and people younger than 45 years old. About 30% of lung cancers in men are adenocarcinomas and 40% in women. In non-smokers cases these percentages are double and adenocarcinoma incidents are more common in Asian populations. Adenocarcinoma takes a long time to grow and may need many years to develop to an invasive cancer. As shown in Figure 2.11, it usually develops near the edges of the lung and may take years before any symptoms are presented. The cells of this type of cancer form glandular patterns. Symptoms include coughing, shortness of breath, bloody sputum, wheezing, and chest pain and they develop very slowly. This type of cancer can appear to be pneumonia or a collapsed lung.
The main cause of adenocarcinoma lung cancer is smoking. Some secondary factors include age, family history, and exposure to smoke or dust. The period a person has been smoking and the amount of cigarettes he smokes play a very important role in the chances to develop adenocarcinoma lung cancer. If he stops smoking then the chances to develop adenocarcinoma start to decrease. [16]

Adenocarcinoma incidents show an upward trend the recent years and according to Lynne Eldridge this might be caused due to the addition of filters to the cigarettes. These filters allow the smoke to reach deeper places of the lung, where adenocarcinoma is developed. [17]

2.3.3 Large-Cell-Carcinoma

Large cell carcinoma, or else large cell lung cancer, is another subtype of NSCLC and it is responsible for about 10% - 15% of lung cancers [13]. This type of cancer usually appears in the central part of the lung, as shown in Figure 2.5. Large cell carcinoma took its name after the large abnormal cells which create - also large - tumours of this type [18]. It grows very fast and may spread to other organs even if the tumour is still small, which make it very hard to be treated [13]. According to the Harvard Health Publications, large cell carcinoma ”is usually discovered at a later stage” [18].

Large cell lung cancer is strongly connected with smoking. Smoking is the number one and the most common cause of large cell lung cancer. [18]
2.4 Small Cell Lung Cancer

According to the College of American Pathologists, Small Cell Lung Cancer (SCLC) is responsible for the 15% of the lung cancer cases. SCLC is also called "oat cell cancer" because when the abnormal cells are looked under the microscope they look like oats. [19] This type of cancer grows very fast and spreads faster than NSCLC, so it is hard to be cured. SCLC is divided into three subtypes of small cell lung cancer:

- Small cell carcinoma (the most common one)
- Mixed small cell / large cell carcinoma
- Combined small cell carcinoma

As can be seen in Figure 2.6, SCLC often shows up inside the bronchi or somewhere near the bronchi. Because of its quick growth, it forms large tumours that can spread immediately to any other parts of the body such as the brain or the liver. The cause of this easy spread is the "constant flow of blood and lymph through the lungs", according to the College of American Pathologists. [19] Usually, SCLC spreads to other parts of the body before it is discovered, which leads to the problem that the cancer cannot be cured only by removing the tumour inside the lung. There are many cases where the SCLC tumours block nearby veins, a problem which is considered as a very serious one. The reason of this situation is that SCLCs usually grow near the chest’s biggest and most important blood vessels. [20]
The primary cause of Small Cell Lung Cancer is smoking as in every other type of lung cancer. Current or past smokers are the 90% of the cases. It is very rare for a person who never smoked before to be diagnosed with SCLC. Other secondary causes of SCLC are the heavy exposure to second-hand smoke, asbestos or radon.

2.5 Lung Carcinoid Tumour

Lung Carcinoid Tumour [21] is the third big category of lung cancer. Cancers of this category are also called ”cancers in slow motion” because they grow very slow so people with lung carcinoid tumours usually live for many years or even a normal lifetime, even if their tumours are malignant. Lung Carcinoid Tumours are created from the uncontrolled development of neuroendocrine cells and they represent approximately the 10% of lung cancers [21]. Lung Carcinoid Tumours start in the bronchi of the lung and do not metastasize to other parts of the body in most of the times, as stated by the Harvard Health Publications. [22]

Lung Carcinoid Tumours are divided into two categories, according to George Schiffman [22]:

- **Typical Carcinoid Lung Tumours:** 9 in 10 lung carcinoid tumours are typical. They grow very slow and there are only a few cases where they metastasize in other parts of the body.
• **Atypical Carcinoid Lung Tumours**: These carcinoid lung tumours are uncommon. They grow faster and more aggressively than the typical ones and there is a higher possibility for them to spread to the patient’s body.

Smoking or any other environmental causes are not related with Lung Carcinoid Tumour. Anyone can develop this type of lung cancer but it is more common in African-American males, as stated in [21]. The cases are equally divided between women and men, usually in the aging range of 45 to 55. [22]

### 2.6 Detection of Lung Cancer and Current Problems

If any symptoms such as those mentioned in Section 2.2 appear, the patient need to see a doctor. If the doctor suspects that the patient may have a lung cancer there are various ways to be diagnosed. These detection mechanisms are explained in the next subsections.

#### 2.6.1 Chest X-ray

Chest X-ray is a safe and effective way to get an image of the inside of the chest [12]. It is the first test which is done to a patient for lung cancer diagnosis. Cancerous tumours are shown as white-grey masses in the X-Ray. [23]

An example of an X-ray with a tumour is shown in Figure 2.7. This X-ray shows that the patient has adenocarcinoma lung cancer. On the left upper side of the picture, which is the right lung, there is a rounded light spot at the level of the second rib. [24]

Unfortunately, X-rays are not a precise diagnose method because there are situations where cancer may be thought of as any other condition or not be seen at all because of its position or because the type of cancer does not show up in X-rays. [23] [25]

As far as the risks of this method are concerned, patients are exposed to radiation which may be harmful to them, although these X-rays used for medical purposes have a very small radiation dose. [26]

#### 2.6.2 CT Scans

CT scan [23] stands for Computerised Tomography Scan and it is usually the next step after the Chest X-Ray (see Subsection 2.6.1). CT scans use X-rays on a computer in order to produce more detailed images of the chest in a 3-dimensional view. Prior to
having a CT scan, the patient is injected a contrast medium in order for the lungs to be shown more clearly on the scan. A CT scan may also be used before a biopsy [27].

An example of a CT scan of a patient with adenocarcinoma lung cancer is shown in Figure 2.8. There is a thickening in the right lung (see arrows) as well as a mass. [28]

In order for the doctor to ask a patient to have a CT scan, there must be a "clear medical benefit", according to NHS. Patients are exposed to radiation which is a bit more than other image tests. However, CT scan is very useful because of the clear images it provides to doctors. [29]
2.6.3 PET scan

The Positron Emission Tomography (PET) Scan is performed after the CT scan (explained in Subsection 2.6.2) in the case that the results of the CT scan reveal a cancer [23]. Unlike CT scan which deals with the anatomy of the human body, the PET scan looks at the functions of the body, how the organs and tissues are working [30]. The PET scan is used to measure the activity of the cells [27] and it can indicate the position of active cancer cells. Prior to a PET scan, the patient is injected with a low-dose radioactive material [23]. PET scan, in contrast with the other detection mechanisms, specifies actively growing tumours [31].

Figure 2.9 shows a CT scan image on top and a combination of the CT scan with the PET scan below. On the CT scan image, a mass is visible in the left lung (upper right of the image). The PET scan shows the cancerous cell activity of the mass and its location precisely.

![CT scan and PET/CT combination image](http://www.columbiasurgery.org/news/si/img/pic_lung_petct.jpg)

PET scans are not available everywhere so if it is necessary to be done the patient needs to find the nearest hospital with a PET scan [27].

2.6.4 Sputum Cytology

Sputum Cytology [32] is the collection of sputum by forcing the patient to cough. The collected sputum sample is examined under a microscope in order to track any abnormal or cancerous cells. The sputum is a mucousy substance which is created in the lung airways. If the patient has lung cancer which is near the airways, then the
sputum of the patient may contain cancerous cells. [33]

Figure 2.10: Normal Pattern Sputum
http://www3.ha.org.hk/qeh/department/path/dept_pathology_dept_histo_sputum.htm

Figure 2.10 shows the sputum of a healthy patient under the microscope. Various types of cells can be identified such as the Bronchial and Squamous cells. Figures 2.11 and 2.12 show the sputum of two patients with Adenocarcinoma and Squamous Cell Carcinoma respectively.

Figure 2.11: Adenocarcinoma Lung Cancer Sputum
Figure 2.12: Squamous Cell Carcinoma Sputum
http://www3.ha.org.hk/qeh/department/path/dept_pathology_dept_histo_sputum.htm

Sputum Cytology is an easy procedure but it is not always an accurate diagnosis mechanism and may miss abnormal or cancerous cells. Furthermore, it is limited to those cases where cancer tumours are growing near the airways. [31]
2.6.5 Bronchoscopy

Bronchoscopy [23] [34] is the procedure where the doctor checks the inside of the patient’s airways and collects a small sample of cells (biopsies) which are removed from the patients lungs. This is done if the CT scan (explained in Subsection 2.6.2) reveals a lung cancer in the central part of the patient’s chest. In order for the sample to be taken, a bronchoscope, a thin tube is used. The bronchoscope gets into the patient’s body through his/her mouth or nose and it is passed down the throat into the airways of the lung. The bronchoscope has a light at its tip and an eye piece to let the doctor see inside. There is also the possibility for the doctor to take photographs of the airways if he/she considers it as necessary.

![Bronchoscopy example](http://lungcarein.com/images/bronchoscopy.jpg)

2.6.6 Fine needle aspiration biopsy

A fine needle aspiration biopsy [35] is an examination done to detect if a tumour is cancerous or non-cancerous. A needle is inserted into the patient’s body through the chest wall in order to take a sample of the tumour. The biopsy is usually guided by a CT scanner or an ultrasound. [31] This procedure is usually performed when the tumour is not reachable through a bronchoscopy (explained in Subsection 2.6.5) [31].

Figure 2.14 shows a lung biopsy with guidance by a CT scan. As can be see there is a mass in the middle of the left lung (right side of the scan) and there is a white straight line which is the needle used for the biopsy.
Figure 2.14: CT Scan guided Lung Biopsy
https://edc2.healthtap.com/ht-staging/user_answer/reference_image/4567/large/Lung_biopsy.jpeg?1386670203
Chapter 3

Background Study - Digital Human Phantom

Chapter 3 explains the Digital Human Phantom (DHP). Section 3.1 explains how the Digital Human Phantom is produced and Section 3.2 explains the structure of the Digital Human Phantom data.

3.1 DHP Production

Digital Human Phantom is a model of the human body which represents the internal structure of a human body. In order to produce such a model there are some steps that need to be followed. These steps are [36]:

1. MRI is used to scan the entire human body from head to the feet. Each MRI image displays a segmentation of the human body, which is perpendicular to the direction of the backbone. The distance between two consecutive MRI scans is usually either 1mm or 2mm. (Figure 3.1)

2. The scanned images are segmented by medical doctors, who identify the different tissues at each pixel of the images. Tissues include bones, fat, muscles, skin and other parts of the body. For our digital human phantom, each pixel’s size in the image is 1mm × 1mm. Hence, each pixel has a number which uniquely identifies the tissue which it shows. For instance, bones are represented by number 10 and left lung is represented by number 85. At this point, a stream of integers without the Cartesian coordinate substitutes the MRI images. The file which has those numbers has a filename which specifies the distance of the MRI scan
phantom from the patient’s head. For example, the scan of Figure 3.2 which is a part of the patient’s chest has a lower number than a scan which is near the feet of the patient.

Figure 3.2 shows an example of a digital human phantom image which has passed from the two steps mentioned before. It is an image of the lungs of a human. The two lungs can be distinguished by their unique colour. The white one (number 85) is the left lung and the dark grey (number 27) is the right lung of the human. The name of the image shows the height of the image from the patient’s head.

Figure 3.2 shows an example of a digital human phantom image which has passed from the two steps mentioned before. It is an image of the lungs of a human. The two lungs can be distinguished by their unique colour. The white one (number 85) is the left lung and the dark grey (number 27) is the right lung of the human. The name of the image shows the height of the image from the patient’s head.
3.2 DHP Images Structure

Section 3.1 explained how the digital human phantom is produced. Section 3.2 describes the structure of the produced images. Because of the huge amount of data in each DHP image, Section 3.2 will use a sample image created specifically for the explanation of the image structure. This image is shown in Figure 3.3.

There are four different colours in the image indicating four different body tissues. The numbers used to create the image are shown in Table 3.1. These numbers are converted to the image in Figure 3.3. Each pixel in the image has 1 mm width and 1 mm height.

![Figure 3.3: DHP image sample](image)

<table>
<thead>
<tr>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>75</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>75</td>
<td>150</td>
<td>255</td>
<td>150</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>75</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3.1: Image numbers
Chapter 4

Research Methodology and Project Planning

Chapter 4 has a descriptive presentation of the methodology used to conduct the literature review. Furthermore, it presents the plan which will be followed for the project evaluation and the project plan followed with the project milestones and deliverables.

4.1 Research Methodology

Research [37] is the "study for knowledge" and "a scientific and systematic search for pertinent information on a specific topic". For research to be done scientifically, a research methodology needs to be followed. This keeps the researcher focused in the right direction without doing irrelevant things which do not add any value to the research. The research methodology followed for the research of this project is explained below:

1. **Formulate the problem:** The problem was firstly initiated and discussed with the supervisor to identify the scope of the problem and the project aims and objectives. The feasibility of a proposed solution was considered in order to be able to start implementing the solution.

2. **Background Study and Literature Survey:** After the problem was understood and the aims and objectives were clearly stated, a survey of the literature was necessary. There was an extensive study about:
   - the digital human phantom
• the structure of the lungs
• the types of lung cancer
• the characteristics of each type of lung cancer, and
• the detection mechanisms

Furthermore, there was a need for a background study on pgm and ppm images and bash scripting in order to be able to start the implementation of the project.

3. **Typical lung cancer shape development**: Based on the literature survey, a typical lung cancer shape will be developed, in order to be used for the next step, the algorithm implementation.

4. **Algorithm implementation**: An algorithm will be implemented to alter the current MRI scans of the lungs and make them cancerous. At first, the text files which include the image’s data need to be converted to valid pgm and ppm images and then select only those images which show parts of the lung. Finally, the algorithm for adding cancer to the healthy digital human phantom will be implemented.

5. **Verification via visualization**: The images produced by the algorithm will be verified by visually comparing them to images from real lung cancer MRI scans.

## 4.2 Evaluation Plan

After the project implementation, the project will be evaluated in order to be sure that the final product meets the requirements, fulfils the aims and objectives and is not out of scope. According to OECD [38], the criteria that need to be taken into consideration for the project evaluation are:

**Relevance**

The resulting product will be checked to ensure that it meets the requirements and is what was expected to be. The produced images will be compared to real MRI scans from patients who have lung cancer in order to be sure that the produced images will reflect the reality.

**Effectiveness**

Ensure that the product meets its objectives and if there are objectives which were not met, the factors which influenced should be stated.
Efficiency
Measure the time needed for the algorithm to function and if that is acceptable or not and check if the implemented algorithm is the most efficient way compared to other alternative options.

Impact
Check what changes this project accomplishes, what benefits it provides to the group and who/what are affected by this project.

4.3 Project Planning
The list given below is an analytical plan of what has been done and what will be done next in order to fulfil the requirements of this MSc Project. Minor changes have been made since the initial report submission in order to synchronize with the current progress of the project.

31/January-7th/February ppm and pgm image understanding
7/February-14th/February Shell scripting learning
14/February-21st/February Survey of the lung cancer
21/February-28th/February Write the initial report
28/February-5th/March Evaluate and correct the initial report
6th/March Submit the initial report (Milestone) (Deliverable)
6/March-23rd/March Survey of the lung cancer, Shell scripting on the ppm file production
23st/March Deliver the shell script on the ppm file production (Deliverable)
24/March-4th/April Shell scripting to identify the location of lung in the DHP
4th/April Deliver the shell script to identify the location of lung in the DHP (Deliverable)
4/April-28/Apr Write the progress report
28/April-8th/May Evaluate and correct the progress report
9th/May Submit the Progress Report (Milestone) (Deliverable)

10/May-6th/June Exam preparation and exams

7/June-29th/June Implement the visualization tool (Deliverable)

30/June-20th/July Create 3D visualization to check the tool

1/August-20th/August Write the dissertation report

21/August-4th/September Evaluate and correct the dissertation report

5th/September Submit the Dissertation (Milestone) (Deliverable)
Chapter 5

Project Progress

Up to this point (May 2014), there was a quite good progress on the project. There was an intensive study of the pgm and ppm images as well as an implementation of shell scripts for the ppm image production and for identification of the location of lung in the DHP. A description of each activity of the progress made so far is given in each Section below.

5.1 Meetings with supervisor

Since the project assignment, there were weekly meetings with the supervisor in order to discuss the progress of the project and the goals for the week. Firstly, there was a discussion about the purpose of this project in order to have the best possible understanding of the aim and the objectives of the MSc thesis. The next thing was to create a plan for the progress of the project.

Each week, the supervisor was reviewing the tasks given before and was assigning new tasks and goals for the next week. She was also answering any questions that may arised. Any problems regarding the progress or the implementation were discussed.

5.2 Understanding of pgm and ppm images

The first task for the fulfilment of this project was to understand the format of pgm and ppm images. All the data had to be converted to these types of images so to be able to do this, their format had to be understood.

The PGM (Portable Gray Map) format is Netpbm’s grayscale graphic image format and it is an array of arbitrary integers [39]. The file containing this image begins with
a number to identify the file type, which in this case is "P5". Next comes the width and height of the image in decimal separated by a whitespace. The next line holds the maximum gray value which is presented in the file. The last information of the file are the numbers which represent the colour of each pixel. An example of a pgm image is shown in Table 5.1 below:

<table>
<thead>
<tr>
<th>P2</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 6</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>0 0 0 0</td>
</tr>
<tr>
<td>0 3 3 3</td>
</tr>
<tr>
<td>0 5 0 0</td>
</tr>
<tr>
<td>0 5 0 3</td>
</tr>
<tr>
<td>0 8 0 3</td>
</tr>
<tr>
<td>0 8 0 0</td>
</tr>
</tbody>
</table>

Table 5.1: PGM file example

The PPM (Portable Pixel Map) format is Netpbm’s color image format and it includes the red, blue and green integer for each pixel of an image [40]. The file containing this image begins with a number to identify the file type, which in this case is "P3". Next comes the width and height of the image in decimal separated by a whitespace. The next line holds the maximum colour value which is presented in the file. The last information of the file are the numbers which represent the colour of each pixel. An example of a ppm image is shown in Table 5.2 below:

<table>
<thead>
<tr>
<th>P3</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 4</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0 15 0 15</td>
</tr>
<tr>
<td>0 0 0 0 15 7 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 15 7 0 0 0</td>
</tr>
<tr>
<td>15 0 15 0 0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>

Table 5.2: PPM file example

5.3 Background study on the position of lung cancer

In order to be able to understand the position of the lung cancer, a background study on the lung cancer was essential. Lung cancer have different types and subtypes. Each
one of them has different characteristics and the cancerous cells and tumours have a
different position in the lung depending on the lung cancer type. The findings of the
background study are discussed in Chapters 2 and 3.

5.4 Shell script for ppm image production

The data from the MRI scans are given in text files which only contains the numbers of
the various pixels in their gray format. The first thing to be done, was to convert those
text files to valid pgm and ppm images. Appendix A lists the code for this script.

The executable is called txt2pgm2ppm.sh and it converts .txt files to .pgm and .ppm
files. It gets as arguments the files to be converted or even whole folders with such text
files. First, the .txt is converted to pgm image by adding the proper header to the text
file and then it is converted to a ppm image using the pgm2ppm executable under the
same folder.

5.5 Shell script to identify the location of lung in the
DHP

The given images include the whole body of the patient. For this specific project, rele-
vant images are only those which contain parts of the lung. In order to select only those
images which represent the lung area, a tool was implemented called getLungImg.sh.
The implementation of this tool is shown in Appendix B.

This tool creates a new folder called "lungimgs" and adds there only those images
which include parts of the lung. To identify those images, the tool checks if the image
contains pixels with the number 27 and/or 85 which represent the right and left lung
respectively.
Chapter 6

Conclusion

Lung Cancer has various types, each one of them having different characteristics which affect the position of the lung tumours and the rapidity of their growth. Therefore, there is not a single algorithm to replace the healthy lung tissues in the digital human phantom.

The next step in this project will be to implement the algorithm for the healthy lung tissues replacement and to evaluate it using images from patients with lung cancer.

This report introduced the lung cancer types and the characteristics of them based on the literature review. Furthermore, there was an explanation of the structure of the digital human phantom. It also provided the project methodology which is followed as well as the project plan. Finally, it described the progress made so far for the fulfilment of the MSc thesis.
Bibliography


Appendix A

Shell Script for ppm and pgm image production

Appendix A lists the bash code for the production of the pgm and ppm images.

```bash
#!/bin/bash

# Functions

function createImage {

    # set pgm image's name
    pgmimg=`echo $1 | sed s/txt$/pgm/`

    # find maximum number in image file
    max=`sort -g -k 1 $1 | tail -1`

    cat $2 > $pgmimg
    echo $max >> $pgmimg
    cat $1 >> $pgmimg

    # set ppm image's name
    ppmimg=`echo $1 | sed s/txt$/ppm/`
    cp $pgmimg test.pgm

    # !/bin/bash
    # ##################################################
    ##
    # Functions#
    ##
    # ##################################################
    function createImage {

        # set pgm image’s name
        pgmimg=`echo $1 | sed s/txt$/pgm/`

        # find maximum number in image file
        max=`sort -g -k 1 $1 | tail -1`

        cat $2 > $pgmimg
        echo $max >> $pgmimg
        cat $1 >> $pgmimg

        # set ppm image’s name
        ppmimg=`echo $1 | sed s/txt$/ppm/`
        cp $pgmimg test.pgm

    }
```

42
# create ppm image
$3/pgm2ppm
mv test.ppm $ppmimg
rm test.pgm
}

# Recursively enters all directories to convert images
function recursiveDirImages {
    cd $1
    array=$(ls)
    for files in $array
done
    # Check if dir or file then convert
    if [ -d $files ]
      then recursiveDirImages $files
    elif [[ $files =~ ^\.[^.]$ ]]
      then createImage $files $2 $3
    else echo "$files is not a valid text file"
      fi
}

#################################################
# # # # # # # # # # # # # # # # # # # # # # # # # # #
# Main Script # # # # # # # # # # # # # # # # # # # # # # # #
# # # # # # # # # # # # # # # # # # # # # # # # # # #
# Create header file
echo "P2" > headerfile
echo "#_pgm_image" >> headerfile
echo "265_490" >> headerfile

headerpath="$PWD/headerfile"
workingdir=$(pwd)

# Create pgm2ppm
# /usr/bin/gfortran -O3 ../bonedata/maxuncompress/bin/point/plan

# For each argument convert to ppm
for i in $*; do

    # Check if dir or file then convert
    if [ -d $i ]; then
        recursiveDirImages $i $headerpath $workingdir
    elif [[ $i =~\.txt$ ]]; then
        createImage $i $headerpath $workingdir
    else echo "$i is not a valid text file"
    fi

done
Appendix B

Shell Script to identify the location of lung in the DHP

Appendix B lists the bash code for the identification of the lung in the digital human phantom images.

```bash
#!/bin/bash

# Functions

# Checks the pgm and ppm images to find lung images
function checkImage {
    right='cat $1 | grep '27' | wc -w'
    left='cat $1 | grep '85' | wc -w'

    # Save lung images tolungimgs directory
    if [[ $right > 0 || $left > 0 ]]
    then
        cp $1 $2/$1
    fi
}
```

# Recursively enters all directories to check images
function recursiveDir {
    cd $1
    array=$(ls)
    for files in $array
    do
        # Check if dir or file then check the image
        if [ -d $files ]
            then recursiveDir $files
        elif [[ $files = ".p[p,g,]m$" ]]
            then checkImage $files $2
        else echo "$files is not a valid pgm or ppm image"
        fi
    done
}

# Filepath to save images containing lungs
filepath="$PWD/lungimgs/"

# For each argument convert to ppm
for i in $*; do
    # Check if dir or file then convert
    if [ -d $i ]
        then recursiveDir $i $filepath
    elif [[ $i = ".p[p,g,]m$" ]]
        then checkImage $i $filepath
    else echo "$i is not a valid pgm or ppm image"
    fi
done