AUTOMATIC ANALYSIS OF 4D LARYNGEAL CT SCANS TO ASSIST DIAGNOSING OF VOICE DISORDERS

Sajini Ruwanthika Gintota Hewavitharanage

ORCID: 0000-0002-4952-1304

Submitted in total fulfilment of the requirements of the

degree of

Master of Philosophy – Engineering (Research)

September 17, 2017

Department of Electrical and Electronic Engineering

The University of Melbourne

'

Abstract

Vocal folds are the two smooth bands of muscles located in larynx just above the trachea. Humans produce voice by vibrating the vocal folds using the air coming from the lungs. This abduction and adduction of vocal folds are controlled by the muscles connected to thyroid cartilage, cricothyroid cartilages and arytenoid cartilages.

When vocal muscles are misused or excessively used, they can be strained or damaged and voice disorders may occur. Furthermore, vocal folds can be damaged and the connecting cartilages and muscles can be affected due to the effect of other illnesses like Parkinson's disease (PD), multiple sclerosis (MS), myasthenia gravis (MG), strokes or tumours.

PD is a neuro-degenerative disease which currently has neither cure nor any pathological tests to detect. The disease progresses very slowly over the years and symptoms appear when approximately 70% of the neuron cells have ceased to function. Usual symptoms are tremors and stiffness in the body muscles which results in difficulty moving most of the body parts externally as well as internally. Consequently, vocal folds and laryngeal muscles get affected and PD patients suffer from vocal impairments. Furthermore, previous studies carried out using laryngoendoscopy, laryngostroboscopy and laryngeal electromyography of PD patients found that those patients have an abnormal phase closure and abnormal laryngeal muscle activity. Moreover, in 2014, a study carried out using a group of early PD patients demonstrated increased glottis area and reduced inter-arytenoid distance in subjects. Therefore, laryngeal measurements could be used as a biomarker for early detection of PD.

However, segmenting the vocal folds region from volumetric laryngeal computed tomography (CT) images is a tedious task, when it is done manually. Manual segmentation schemes require lot of expert knowledge and time, and often provide poor objective and reproducible results.

In this project, we hope to develop a novel automated algorithm to segment the vocal folds region and measure the laryngeal parameters. This thesis consists of two major parts; first it proposes a fully automated segmentation method for segmenting the rima glottidis from 3D laryngeal CT scans and generates the time series for rima glottidis areas, which in future can be used to develop an automated diagnosis tool for voice disorders. The gray-level difference features are learnt through a support vector machine classifier and several post processing algorithms are introduced to refine the final segmentation result.

Second, a fully automated method to estimate the vocal plane position in a 3D laryngeal CT volume using computer vision algorithms and techniques is proposed. Vocal plane position is identified using anatomical markers like thyroid cartilage and vertebral bones and these markers are segmented using gray-scale and edge-based features.

The experiments are conducted using a private data set from the Movement Disorder Clinic at Monash Medical Centre The detailed implementations of the two methods including feature extraction, kernel selection, post processing and validation are explained in this thesis.
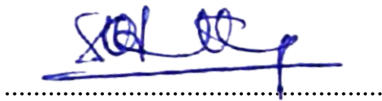
Declaration

This thesis comprises only my original work undertaken towards the degree of Master of Philosophy (Research) except where indicated and has not been submitted in whole or part for a degree in any university.

To the best of knowledge, this thesis does not contain any material previously published by any other person.

Due acknowledgement has been made in the text to all other material used.

This thesis is fewer than 50,000 words in length, exclusive of tables, maps, bibliographies and appendices or that the thesis is [36,478] words as approved by the Research Higher Degrees Committee.

....

Sajini Hewavitharanage

September 2017

Dedication

To my parents

The reason of what I become today

To my husband Chalaka

Without whom this does not exist

To my daughter Kimaya & my son Kavin

For your love

Preface

This thesis has been done at the Department of Electrical and Electronic Engineering in The University of Melbourne. Two conference papers were published from the main parts of this work, Chapter 3 and 4.

Chapter 3 is based on the following publication:

- S. Hewavitharanage, J. , D. Thyagarajan, K. Lau and M. Palaniswami, "Automatic segmentation of the rima glottidis in 4D laryngeal CT scans in Parkinson's disease," *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan, 2015, pp. 739–742.

Chapter 4 is based on the following publication:

- S. Hewavitharanage, J. Gubbi, D. Thyagarajan, K. Lau and M. Palaniswami, "Estimation of vocal fold plane in 3D CT images for diagnosis of vocal fold abnormalities," *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan, 2015, pp. 3105–3108.

As the first author of both papers, I conducted the research, development, implementation, validation, writing of the publications and editing of the comments from co-authors and referees.

The third party copyright material included in this thesis are listed as follows.

Citation Information for the Third party copyright material	Location of item in thesis	Permission granted Y/N
Parkinson's disease related vocal folds change images (a), (b) and (c)	p.2	Y
Parts of the laryngeal cavity	p.6	Y
Different views of laryngeal anatomy (a), (b) and (c)	p.7	Y
Thyroid cartilage anatomy	p.8	Y
Cricoid cartilage anatomy	p.9	Y
Arytenoid cartilages anatomy	p.9	Y
Vocal folds anatomy	p.11	Y
Layers of the vocal folds	p.12	Y
Different cross sectional views of human larynx (a), (b) and (c)	p.91	Y
Movement of vocal folds and connecting cartilages during speech process (a) and (b)	p.92	Y
Changing of airway width with the axial anatomy (b)	p.105	Y

Acknowledgements

This work which I have carried out during my Master's degree would not have been possible without the support of many colleagues, friends and family. First I would like to convey my sincere gratitude to my principle supervisor Professor Marimuthu Palaniswami for his constant guidance, encouragement and support. Thank you for accepting me as a Master student to the laboratory, believing in my abilities and supporting me through out the candidature to overcome all the hurdles.

Also I would like to thank my co-supervisor Dr. Jayavardhana Gubbi for the time and effort you devoted to my project and publications. I greatly appreciate the technical and academic guidance you always provided me and the support given during difficult situations.

Furthermore, I would like to extend my gratitude to my advisory committee chair, Associate Professor Tansu Alpcan for his time spent in my progress meetings. Your advice and comments improved the quality of the thesis immensely. Many thanks for Dr. Sutharshan Rajasegarar as well for attending as an advisory committee member and providing your valuable comments. My special thanks go to Professor Dominic Thyagarajan, the director of Neurology and Professor of Neuroscience at Monash Medical Centre, for collaborating with me, and providing ideas and the CT data required. Also, I would like to mention the head of Thoracic Imaging and the CT Unit at Monash Medical Centre, Clinical Professor Kenneth Lau, for his contribution given for publications and to clarify the problems in data.

I am grateful to my colleagues, Fuchun Hsu and Aravinda Sridhara Rao, for sharing

their machine learning with me to overcome difficulties during the project. Moreover, I would like to thank my other laboratory mates, for their support during my years in the laboratory with the inspiring conversations and for making it a fun work environment. Also, many thanks for the staff in the Electrical and Electronics Engineering Department and Research Higher Degree Department at the University of Melbourne for their valuable assistance during my candidature.

Finally, I would like to remember all my friends, especially who are from the University of Melbourne, for their endless support during difficult times. My post-graduate career is greatly enriched by your friendship.

Last but not least, my great gratitude from the bottom of my heart goes to my family for their love and support. My parents and my husband, Chalaka, were always there for me when I needed them and never stopped believing in my success even when I did not. Thank you for looking after Kimaya and Kavin and giving me the freedom to concentrate on my work. If not for them, this thesis would not have been possible by any means. This work is dedicated to them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1   Motivation

Vocal folds are responsible for producing the sound of human voice. The air expelled from lungs during the respiratory cycle is modulated by the vocal folds to produce voice. The voice of a person depends on the size and shape of the vocal folds, strength of airflow, coordination and strength of laryngeal muscles and cartilages and size and shape of laryngeal cavity (nose, throat and mouth). Damage to any of these laryngeal structures may cause voice disorders. Voice disorder is defined as abnormal voice quality of an individual that interferes with his or her daily activities.

Voice disorders can be classified in–to two categories: organic and functional. Organic disorders occur when vocal folds and surrounding laryngeal structures are altered or damaged. Functional disorders occur when vocal folds are abused or inappropriately used. In this case the physiology of the laryngeal cavity is normal, but the laryngeal structures may be strained or exhausted. Organic disorders may be structural, where the vocal folds or laryngeal cartilages are abnormal due to a pathology or aging, or neurogenic, where laryngeal function has been affected due to some neurological disorder. However, most of the time these types are not mutually exclusive. For an example, strained vocal folds may be a result of misusing the vocal folds which is a functional disorder, however, repeated use of strained vocal folds may cause trauma and structural damage as well.

Neurological vocal abnormalities have a wide range of causes, including local laryngeal disorders in the head or neck region or neuromuscular disorders which occur in the whole body like Parkinson's disease (PD), multiple sclerosis(MS), myasthenia gravis (MG) or strokes. Most often vocal abnormalities are a part of neurological disease progression. The underline neurological condition may affect the motor control of larynx, weaken or cause atrophy of the laryngeal muscles, or cause the loss of stamina and stability of vocal folds (Figure 1.1).

(a) Breathing position of the larynx. Vocal folds are remarkably degenerated and bowed.

(b) Phonatory position. Left vocal fold bowed and the left vocal process overlaps the right.

(c) Vocal folds coming in to phonatory position. Due to degeneration, contours of the vocal processes are visible.
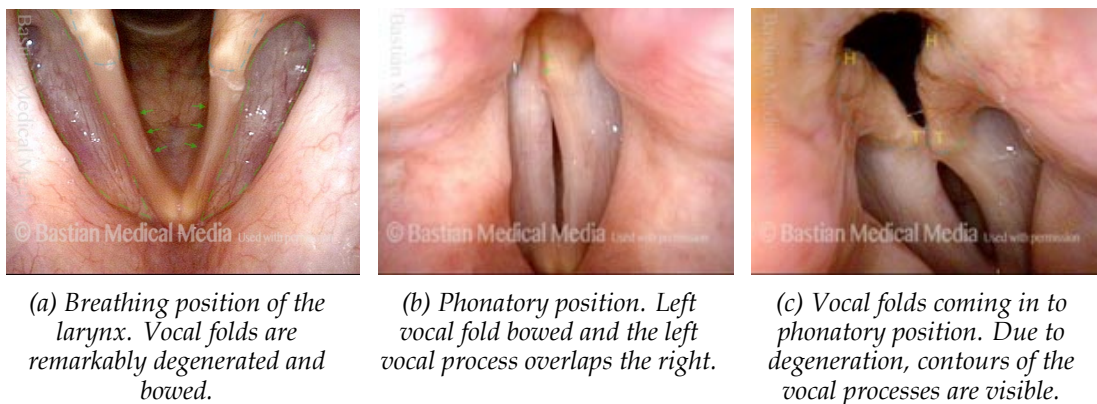
Figure 1.1: Parkinson's disease related vocal folds change[1]

Especially in these neuromuscular disorders (PD, MS and MG), voice problems are one of the earliest symptoms, while other symptoms are often late to appear and hard to distinguish as several other disorders can mimic the same symptoms as well. Diagnosis of these neurological disorders is through examination of the patient history and closely monitoring the disease progression. There is no specific cure for these disorders and medication can be used to control the symptoms and improve the quality of life. These disorders progress very slowly over a long time and therefore most of the time the diagnosis is made at the severe stage. This makes the controlling the disease more difficult.

PD is the second most prevalent neurological disorder after dementia in Australia. At the moment, 10 million people in worldwide are suffering from PD. In Australia, approximately 80,000 individuals have the disease. Twenty percent of that number are from the working age population. The estimated cost of the disease to the Australian economy is $775 million in 2011-2012. These numbers are growing in each year and over the last 6

---

[1]http://laryngopedia.com/parkinsons-disease-related-voice-change/

years it has been found that the number of PD patients has increased by 17% and cost over 48%. At this rate, PD is more common than prostate cancer and bowel cancer.

MS is also another common disease though mostly seen in young adults. MS is unpredictable in nature and therefore often misdiagnosed until the central nervous system is severely damaged. Approximately 2.5 million people worldwide are suffering from MS. Australia has over 23,000 patients living with this condition and the majority are between the ages of 20 and 40. The nature of this disease makes it difficult for these people to be in the labour force.

MG , however, though it is not so common, approximately 1.2 out of 10,000 people are suffering in Australia. This is more common in women than men before the age of 40 years. However, more men get MG after the age of 50 years. Although, this is normally not life–threatening the complication may be serious when the neck muscles are affected.

Slurred speech is the first sign of a stroke. Stroke often results in permanent or temporary paralysis. Depending on the area affected in the brain, other symptoms like swallowing or vision problems may occur. According to the statistics reported in 2012, approximately 2% of Australian had had a stroke at some point in their life. With the advances of healthcare, stroke mortality rates have fallen by 70% from 1979 to 2010 and continue to fall. However, timely diagnosis is crucial in preventing and successfully treating a stroke, and determines the quality of life after a stroke event.

As voice disorders are the first sign of these neuromuscular diseases, it is a good biomarker to incorporate in the clinical diagnosis procedure of these disorders. Due to the loss of motor control of laryngeal cartilages and vocal folds during these diseases, the mechanical parameters during the vocalisation are expected to change compared to a normal larynx.

The current method of examining laryngeal parameters is computed tomography (CT) which allows full 3D view of the larynx. However, CT imaging produces a large amount of digital data which is not feasible to manually analyse. It takes a large amount of spe-

cialised labour, time and effort. Furthermore, due to inter- and intra-observer variabilities, obtaining a reproducible result is also questionable. Due to this reason, an automated method of analysing vocal disorders is essential. However, no known automated method has been developed in present literature to analyse CT images to measure vocal parameters.

1.2 Goals

The goal of this research is to design, implement and evaluate an automated tool to measure vocal fold parameters. Laryngeal CT images can be used to analyse the vocal anatomy and an automated method can be developed to segment the vocal fold structures to measure the parameters. These measurements can be used with patients medical history to achieve more accurate and faster diagnosis. While this tool can be used to diagnose any vocal abnormality, our main objective is to use this to diagnose vocal abnormalities associated with neurological disorders. As current methods for detecting neurological disorders like PD, MS and MG take a long time, we are hoping to shorten this diagnosis time by providing a faster way to analyse vocal fold abnormalities associated with these diseases.

The performance of the tool depends mainly on the quality of the CT images that can be obtained. However, to analyse the dynamics of the vocal folds, the CT images need to be taken while speaking. Due to this reason, some amount of blurring is expected and the tool should be able to cope with this. Finally, the tool should be operated with a small false positives rate as the outcome of the analysis should not diagnose a healthy person as a patient. But, due to the nature of these neurological disorders, suspected patients will go through a continuous monitoring process to validate the diagnosis, and therefore having a less false negatives rate is always desirable over less false positives rate. Hence, the performance of the automated tool will be a trade-off between false positives and false negatives rates.

1.3 Research Background

This section explains the background knowledge of this research. First, anatomy of the human larynx is described to understand the vocal mechanism. Second, laryngeal CT imaging technology is briefly explained. Third, causes of vocal disorders are presented with the main focus on prominent neuromuscular disorders. Finally, current diagnosing methods of laryngeal disorders are described and challenges in those methods and the requirement of computer-aided methods are explained.

1.3.1 Anatomy of Human Larynx

The larynx is the main part of the human neck that is responsible for voice production and is normally called the voice box. It is located at the anterior side of the neck just above the trachea and the anterior side of the lower pharynx. Normally, an adult male larynx is $5\ cm$ in length and adult females have shorter lengths [1]. The larynx is located at the level of $C5$ vertebra in infants and descend to the level of $C6 - C7$ vertebra after the age of 2 years while maintaining the relationship between the internal and external parts.

The main function of the larynx is to use laryngeal muscles to temporarily block the airway by halting the breathing during swallowing, so that the airway will be protected from food and other foreign substances. Apart from this, the larynx houses the vocal folds which are responsible for producing sound. Various laryngeal muscles and cartilages act together to produce phonation and control the pitch and volume. Furthermore, the larynx acts as a sensory organ and is also responsible for coughing.

The tube-shaped laryngeal cavity is divided into three main parts, the laryngeal vestibule (supra-glottic space), the middle part and the infra-glottic space (Figure 1.2). The vestibule starts from the laryngeal inlet, which opens to the pharynx, and ends at the vestibular folds. The part between the vestibular folds to the vocal folds is called the middle or the voice box. The inferior part from the vocal folds to the opening of the trachea is called the infra-glottic space. On the either side of the middle larynx, there are two saccules formed by the mucosal membrane along the entire length of the middle part. These tubu-

lar troughs are called the laryngeal ventricles and are responsible for causing the noisy breathing, snoring, nasal congestion and coughing. The triangular shaped opening between the vestibular folds or false vocal folds are called rima vestibule and the similar shaped opening between the true vocal folds is called the rima glottidis. The rima glottidis is the narrowest part of the laryngeal cavity in human adults [2].
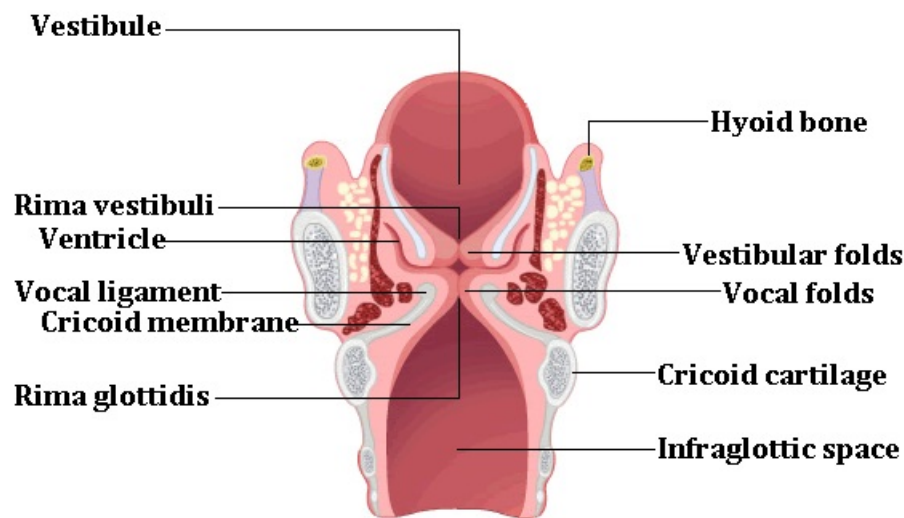
Figure 1.2: Parts of the laryngeal cavity
(Reproduced with permission from the Royal College of Anaesthetists)[2]

The larynx consists of few cartilages and muscles. Hyoid bone, which situated in the anterior mid-line of the human neck and supports the tongue, is not a part of the structure of the larynx. However, this bone provides the muscle attachments to support the motion of the laryngeal cartilages. Among the cartilages thyroid cartilage, cricoid cartilage and epiglottis are the larger unpaired cartilages and the larynx also includes another three paired cartilages, arytenoid cartilages, corniculate cartilages and cuneiform cartilages. Between these cartilages there are number of internal muscles which aid the function of larynx as can be seen in Figure 1.3.

---

[2]http://e-safe-anaesthesia.org/sessions/14_01/d/ANAE_Session/520/tab_720.html
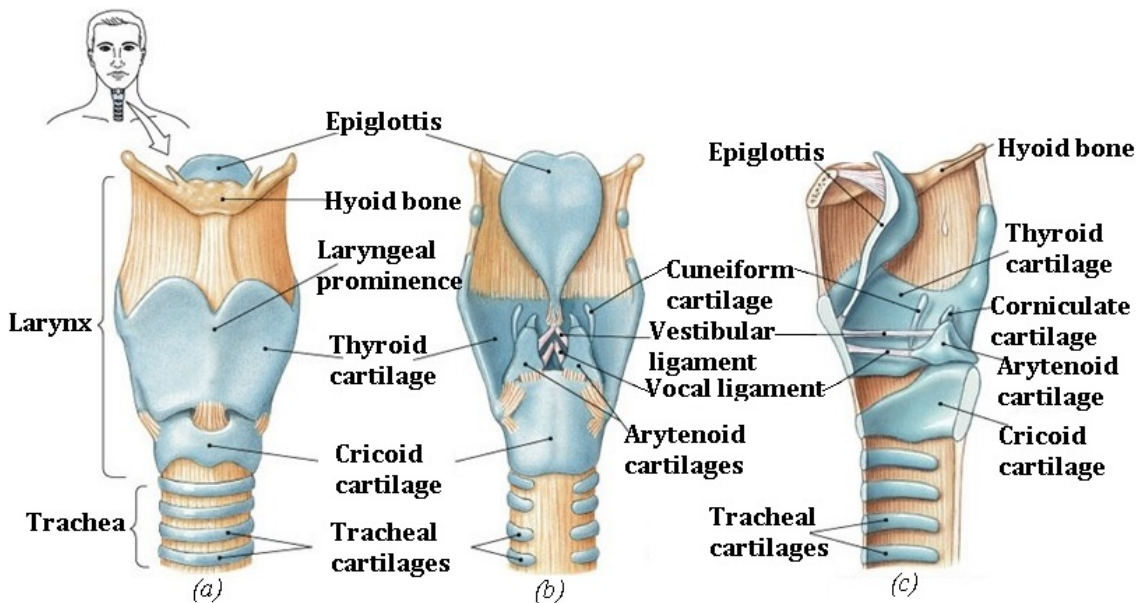
Figure 1.3: Different views of laryngeal anatomy[3] (a) Anterior view (b) Posterior view (c) Mid-sagittal view

Thyroid cartilage: This is the largest cartilage in the human neck skeleton and consists of two left and right lamina which are disconnected posteriorly and cover each lateral side of the trachea (Figure 1.4). The two laminae meet on the anterior mid line of the neck forming an acute angle which known as the laryngeal prominence. The laryngeal prominence is more apparent in men than women. It is approximately $90°$ in men and $120°$ in women.

Superior thyroid notch is the 'V'-shaped joint on the anterior mid-line above the laryngeal prominence. Below the laryngeal prominence on the same mid-line is the inferior thyroid notch which is not very distinct compared to the superior one. The posterior sides of each lamina are extended and form a pair of superior horns on the top and a pair of inferior horns on the bottom. Inferior horns and the inferior edge of the thyroid cartilage are connected to the cricoid cartilage by cricothyroid membrane and cricothyroid muscles. These inferior horns move with the cricoid cartilage during articulation. Superior edge of the thyroid cartilage is connected to the hyoid bone by thyrohyoid membrane.

---

[3]https://www.slideshare.net/drvinaybhat/anatomy-of-larynx-and-trachea-final
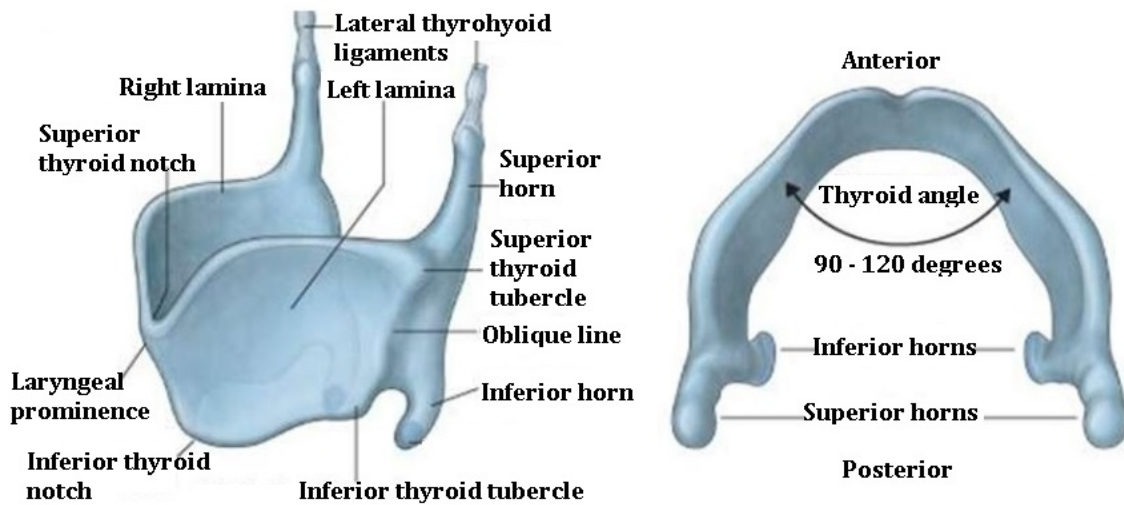
Figure 1.4: Thyroid cartilage anatomy[4]

Epiglottis: This is the leaf-shaped flap-like elastic cartilage which acts like a protective barrier for trachea against food aspiration. It is covered with a mucous membrane and attached to the inner side of the thyroid cartilage around the middle of the inferior notch and laryngeal prominence. During breathing, it remains open to allow air to enter the larynx, but during swallowing, falls down to close the entrance to the trachea to stop food entering the larynx, and goes down the oesophagus.

Cricoid cartilage: This is the ring-shaped cartilage located in the inferior part of the larynx in the level of C6 vertebra (Figure 1.5). Unlike other cartilages, cricoid cartilage has a complete ring shape around the trachea. The posterior part is broader and called the lamina, whereas the inferior part is narrower and called the arc of the cartilage. Cricoid cartilage has two superior facets which are connected to the arytenoid cartilages with cricoarytenoid ligament and articulates with them during speech. Similarly, the inferior lateral facets are articulated with the inferior horns of the thyroid cartilages during speech. The inferior border of the cricoid cartilage is connected to the first tracheal ring with the cricotracheal ligament.

---

[4]https://www.slideshare.net/karthickanatomist/anatomy-of-larynx-31533041!
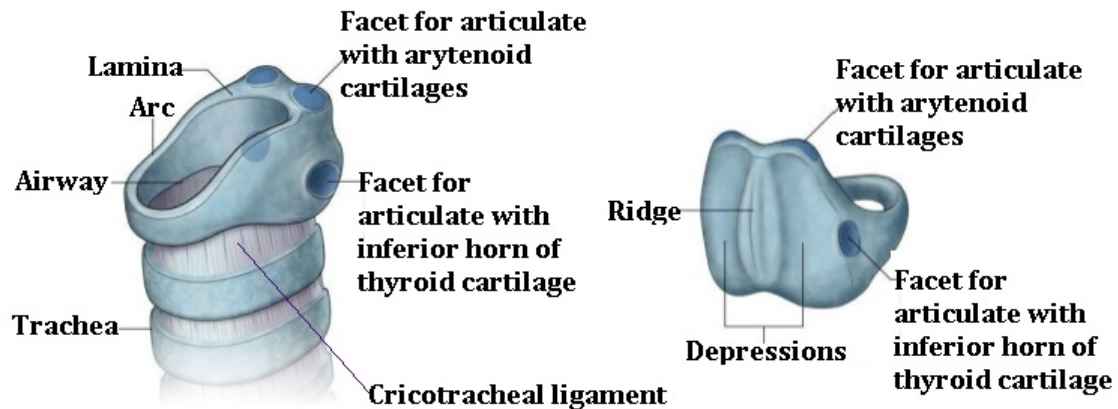
Figure 1.5: Cricoid cartilage anatomy[4]

Arytenoid cartilages: These are two pyramidal shaped cartilages located in the posterior part of the larynx (Figure 1.6). These cartilages have three surfaces, a base and an apex, where the bases are connected to the superior facets of the cricoid cartilage. Furthermore, each arytenoid cartilage has a vocal process which is connected to the vocal ligament and a muscular process which is attached to the posterior lateral cricoarytenoid ligament. Moreover, the anterior lateral surface gives the attachment to the false vocal folds or vestibular ligaments.
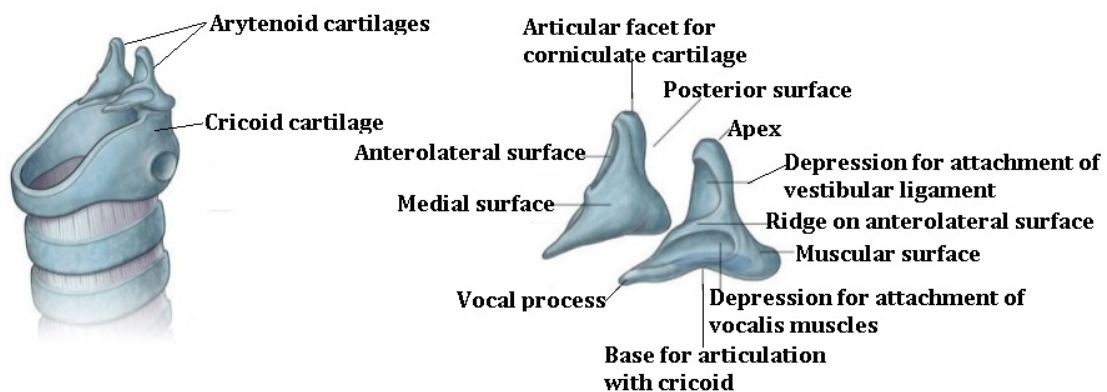
Figure 1.6: Arytenoid cartilages anatomy[4]

Apices of the arytenoid cartilages are articulated with the corniculate cartilages, which are small conical shaped elastic cartilages that helps to extend the arytenoid cartilages posteriorly and medially. There is another wedge-shaped pair of cartilages located an-

---

[4]https://www.slideshare.net/karthickanatomist/anatomy-of-larynx-31533041

terior to the corniculate cartilages and articulate with the arytenoids called cuneiform cartilages. These two yellow elastic cartilages support the vocal folds and lateral part of epiglottis.

Vestibular folds: These are inferiorly attached to the inner aspect of the thyroid carti-lage just below the place where the epiglottis is attached. They are two thick folds made of mucosal membrane and posteriorly attached to the anterolateral surfaces of the ary-tenoid cartilages just above the vocal process. Vestibular folds are known as false vocal folds and not directly associated with sound production. However, they help to lubricate the true vocal folds with mucous and aid to keep food away from the lower part of the airway. Also, false vocal folds overlap with the true vocal folds during the coughing or sneezing to protect true vocal folds from internal pressure.

Vocal folds: These are also two membranous folds located just below the vestibular folds and inferiorly attached to the anterior commissure (AC) of the thyroid cartilage (Figure 1.7). AC is located on the mid-line of the inner aspect of the thyroid cartilage between the laryngeal prominence and inferior thyroid notch. The posterior sides of the two vocal folds are attached to the vocal processes of the arytenoid cartilages. Normally, the location of the vocal process of the arytenoid cartilage is lower than the AC. There-fore, the vocal folds have a downward slope in the posterior direction in the majority of people [3]. Vocal folds produce sound by vibrating the two folds to modulate the airflow coming from the trachea during exhalation.
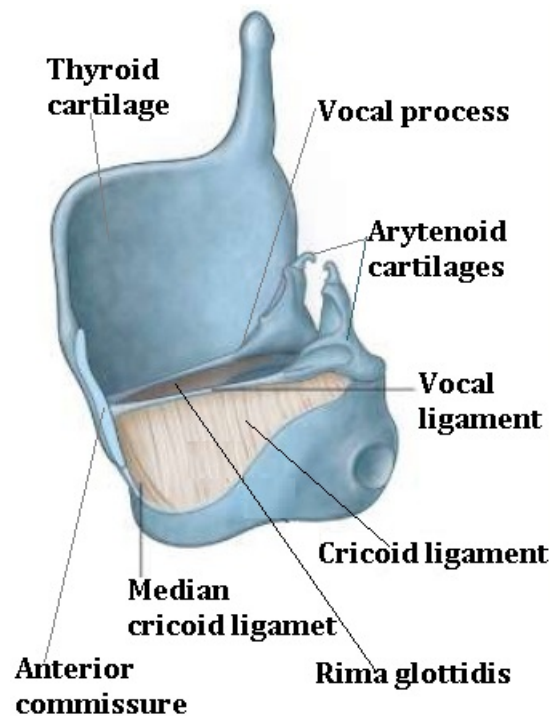
Figure 1.7: Vocal folds anatomy[4]

Normal adult human vocal folds are $12 - 24\ mm$ in length and and $3 - 5\ mm$ in thickness [4] and consists of three main layers of tissues: epithelium, lamina propria and vocalis muscle (Figure 1.8). Epithelium is a very thin lining of tissue that holds the shape of the folds. The lamina propria has three layers called the superficial layer, intermediate layer and deep layer. Lamina propria and vocalis muscle are made from layers of densely packed elastic and collagenous fibres and serve the important functions of the phonatory process.

1.3.2 Laryngeal CT

CT uses X-ray to take images of thin slices at different angles through the body part under investigation, so that the user can view the inside of the body part without cutting it. Older CT scanners were very slow and only took one slice at a time whereas modern scanners with the multi-detector technology can scan multiple slices, ranging from 4 to 640 at a time (Figure 1.9). After taking fine slices of images though the body, radiologists
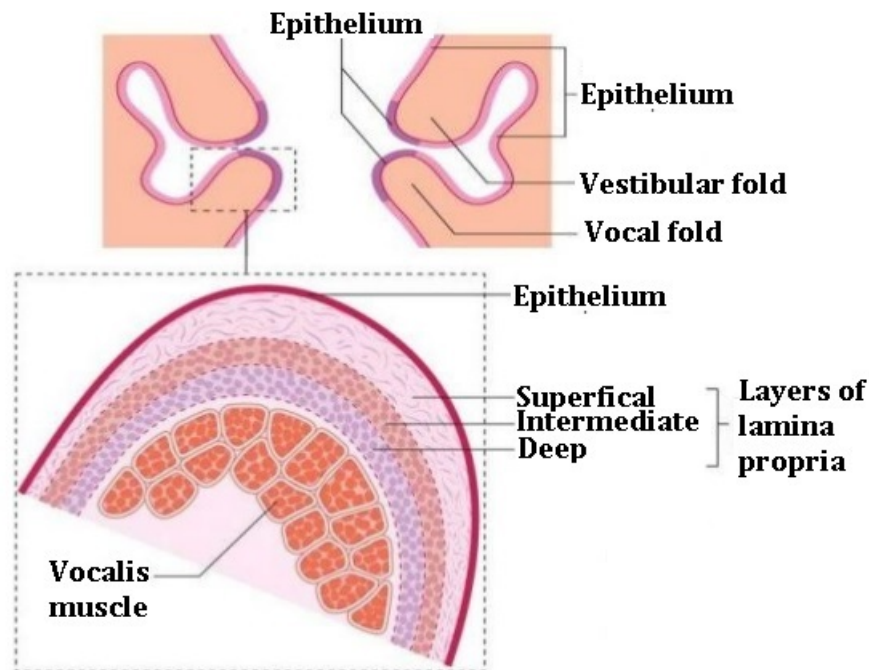
Figure 1.8: Layers of the vocal folds[5]

can combine all the slices and reconstruct 3D image volume of the body. Furthermore, they can slice the image volume in different directions of interest before investigation (Figure 1.10). These images provide very detailed view of the body hence allow very accurate clinical diagnosis.

Endoscopy is the traditional method of observation of human larynx for many years. It provides easy visualisation of the inner surfaces of the larynx and better diagnosing of mucosal and sub-mucosal lesions. However, observing deeper structure is difficult in endoscopy. Magnetic resonance imaging (MRI) or CT imaging provides more accurate non-invasive visualisation of laryngeal structures.

---

[6]http://www.ct-tekijyuku.net/basic/evolution/evolution002.html
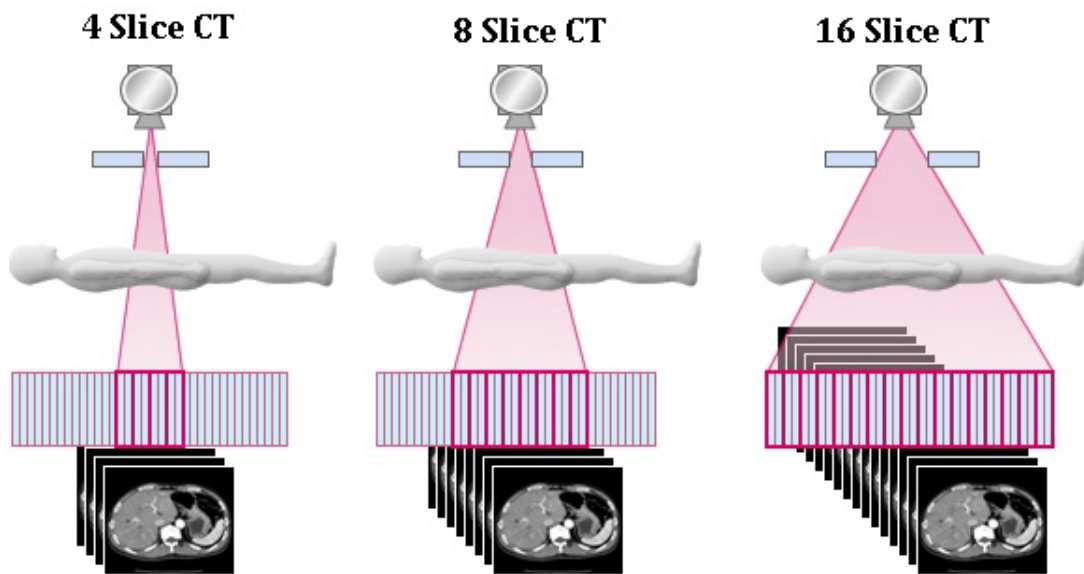
Figure 1.9: Modern multi-detector CT taking multiple slices of images of a body part at once[6]
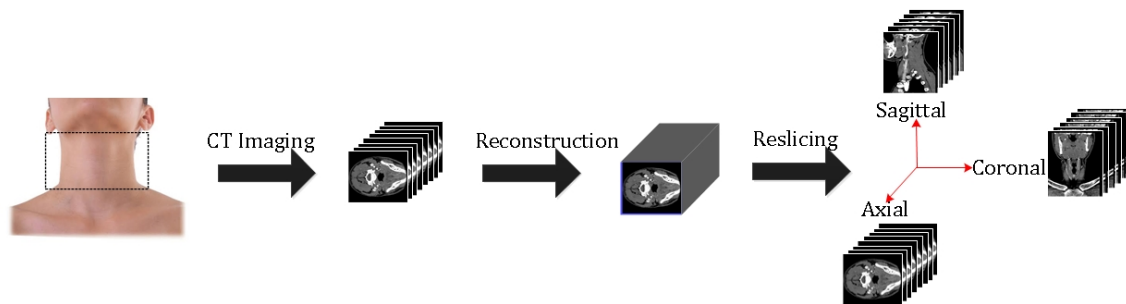
Figure 1.10: Human laryngeal CT imaging, reconstruction and reslicing process

For clearer laryngeal imaging the patient often needs to hold his breath for sometime. Due to this reason, CT imaging is more popular in general laryngeal imaging over MRI as modern CT imaging only takes a few seconds to image the entire larynx. While MRI provides radiation-free imaging, CT is more useful for patients who cannot stay still for examination. Also, modern CT imaging provides 3D images of the larynx allowing the medical professionals to study different cross-sectional views (axial, sagittal and coronal) of the larynx.

In a CT image different body parts like bones, cartilages, blood vessels and soft tissues shows different intensities due to the different absorption rates of X-ray energy. Therefore, these structures can be easily identified in the image. Bones have the highest

absorption rate hence present with the highest intensity whereas soft tissues have low absorption rates and show low intensities.

Therefore, in a laryngeal CT image, the hyoid bone and spinal column always have the highest intensity values and can be recognised very clearly. The next most recognisable structure is the thyroid cartilage. However, the arytenoid cartilages and cricoid cartilages are poorly visible due to their very similar absorption rates compared to the surrounding soft tissues. Air absorbs almost all X-ray energy passing through and appears dark in images. Therefore, structures like rima glottidis, airway and lungs are clearly recognisable in a laryngeal CT image (Figure 1.11).

(a) Mid-sagittal plane   *(b) Axial plane at rima glottidis level*
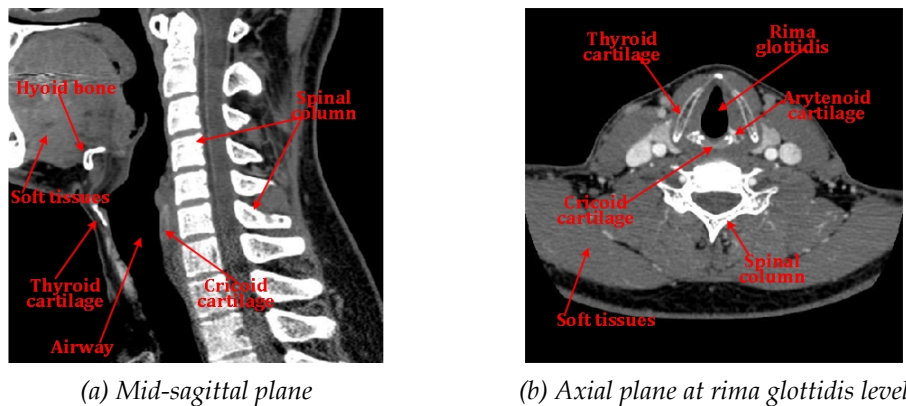
Figure 1.11: Different structures on a laryngeal CT image

1.3.3   Voice Disorders and Causes

Vocal folds are responsible for producing voice. The phonatory process occurs when air is expelled from the lungs through the glottis, creating a pressure drop across the larynx. Normally, the vocal folds remain closed until this laryngeal pressure is sufficient enough to draw them apart. When they are opened, air escapes through the glottis reducing the pressure enough for the muscle tension to pull the folds back together again. This abduction and adduction of vocal folds are controlled by the muscles connected to thyroid cartilage, cricothyroid cartilages and arytenoid cartilages. Vocal folds vibrate and modulate this airflow producing basic sounds which is called phonation.

Excessive or incorrect usage may damage or strain the vocal muscles and can pro-

duce various speech-related problems. Furthermore, various disease can also interrupt the nerve inputs to the laryngeal muscles and can cause vocal folds paralysis. When vocal folds paralysis occurs the movement of laryngeal muscles and cartilages may be completely stopped or weaken. Moreover, various vocal folds pathologies may also result in speech problems. These vocal disorders can happen at any age and change the pitch, loudness or quality of voice affecting people to different degrees. However, it is important to detect these abnormalities and causes early and treat them to save the quality of life of individuals.

Neurological voice disorders

Neurological diseases such as PD, MS, MG or stroke may result abnormal control or coordination problems in laryngeal muscles and cause various voice-related problems.

Parkinson's disease (PD):   PD is a degenerative neurological disorder, which progressively deteriorates neuron cells and causes deficiency of dopamine, a chemical neurotransmitter required to control body movements. The disease progresses very slowly and normally the symptoms appear when 70% of the neuron cells have been affected. Usual symptoms are tremors, stiffness in muscles and slowness of the movements. These symptoms gradually become severe with disease progression thus reducing the quality of life of the affected individual and making the person physically and mentally dependent and socially isolated. Currently, there is no treatment available to cure PD. Furthermore, there are no pathological tests or known biomarkers to diagnose PD, and the diagnosis is solely based on the history of degeneration and the degree of tremors. Due to the complex nature of the disease and the diagnostic procedure, PD is often diagnosed in its severe stage and hence compromising the effectiveness of the treatment and rehabilitation. Therefore, early diagnosis of PD is crucial to slow down or control the disease progression.

These tremors and stiffness in muscles can be presented both internally and externally. Therefore, vocal folds and other speech articulatory muscles in larynx and face are also

affected by the disease. Due to this reason, speech and voice disorders are more common among PD patients. Usually, 72%-89% patients with PD are suffering from voice-related disorders [5], [6].

When these muscles and vocal folds are impaired due to PD, movement of laryngeal cartilages are reduced and this results an incomplete glottal closure. Therefore, the gap between the vocal folds (rima glottidis) is also affected [7]. Due to these reasons, the voice of a patient with PD can be weakened as low as 10 decibels compared to an average speaker. These voice-related pathologies can be characterised as hypophonia(reduced speech volume or intensity), dysphonia (poor voice quality), hypokinetic articulation (reduced range of articulatory movements), dysprosodia (reduced pitch), rush (tendency of speech articulation to festinate), hesitant and disfluency in speech [8], [9], [10].

Current research has shown that these vocal impairments can be one of the earliest indicators for the onset of the illness [11]. Even in early PD, although the speech abnormalities may not be noticeable to the listeners, it has been found out that PD patients have varying pitch/loudness and breathiness in their voice [5]. Therefore, these voice disorders could be a good biomarker for early detection of PD.

Multiple sclerosis (MS):   MS is a very common disease that affects the central nerve system of brain. When the protective cover (myelin) around the nerves is damaged, messages sent through the nerves are distorted. Damage can occur at multiple locations hence the disease is called 'Multiple sclerosis'. The damage to the nervous system can be different from individual to individual, and therefore, the symptoms can also differ.

The main symptom of MS is loss of muscular control which causes weakness in coordination or functioning of legs, arms and other muscles in the body. Furthermore, MS patients suffer from fatigue, vision problems, dizziness and other neurological symptoms such as depression, loss of memory and cognitive problems. Due to the lack of coordination in muscles voice-related disorders are more common among MS patients. Hoarseness and difficulties in controlling the pitch and volume of speech can be seen more often. Also, MS patients suffer from speech problems like slowness and hesitancy

in speaking.

Normally, the disease onsets in younger ages as $20 - 50$ years and will take decades to progress. Comparatively, Caucasian population and females are more prone to MS. As the disease progresses over a long period, the nerve dysfunction can be increased or decreased. Therefore, the patient may have normal behaviour in between the abnormal periods. Due to this reason diagnosing MS is very difficult. Long-term monitoring of the symptoms like muscular weaknesses and voice-related problems along with CT or MRI images of the brain help medical professionals to diagnose MS.

There is no cure available for MS at the moment. However, treatments are available for managing the symptoms and controlling the progression of the disease. Also, speech therapies are used to control the voice-related disorders and increase the quality of life of the patient.

Myasthenia gravis (MG): MG is an autoimmune disease which the antibodies produced by the immune system of the body attacks the acetylcholine receptors at neuro-muscular junctions. Acetylcholine receptors receive the acetylcholine protein that controls the muscle contractions from nerves. When these receptors are attacked, the muscular strength deteriorates and the patients are more susceptible to fatigue. First, the tiredness due to weak muscles may improve after a rest. But gradually, when the disease progresses the tiredness and fatigue will become worse.

Normally, at early MG stages the patients may experience trouble in their vision and eye movements. Then facial expressions and muscles in neck and mouth are also affected. Gradually, moving legs, arms or head may become difficult and even breathing can be affected. When neck muscles are affected, patients will experience voice- and speech-related problems such as dysarthria, hypophonia, hoarseness and difficulty in controlling the pitch.

MG is more common in females than in males and typically onsets between 20 and 40 years of age. Diagnosis is based on the blood tests to detect antibodies and observing

the patient history. However, as most of the symptoms of MG are common with other neurological disorders, clear diagnosing is always difficult. Medications are available to successfully control the symptoms like muscle weakness and voice-related problems.

Strokes: Strokes are caused by an interruption of blood vessels carrying blood to the brain and often cause temporary or permanent paralysis of the body. Due to a stroke one side of the vocal folds muscles can be paralysed and vocal dysfunction can occur. Depending on the region affected in the brain, many patients may experience dysarthria, hoarseness and slurred speech and some patients may lose the ability to speak completely. Other symptoms are poor coordination of arms and legs, numbness and loss of sensation in body parts and vision problems.

Physical examination and patient history may help to diagnose a stroke while a MRI of the brain will tell which area of the brain is affected. Early diagnosis will help to treat strokes successfully with clot dissolving drugs. Also, rehabilitation with speech therapies are often carried out to improve the quality of life of the patient.

Other disorders

Cancerous or non-cancerous tumours can grow in vocal muscles, cartilages and nerves and can interrupt the normal function of the larynx. This can cause various voice-related problems including vocal folds paralysis. Sometimes, tumours in the brain may also interrupt the nerve inputs in the larynx and cause vocal folds problems.

Inflammation caused by various viral infections may cause damage to the nerves in the larynx and may interrupt with the normal function of voice production. Moreover, surgery on the thyroid glands, esophagus, neck or chest may damage the nerves in the larynx and can cause vocal cord paralysis. Similarly, injuries to the neck or chest area also can cause nerve damage in the voice box and cause voice-related problems.

1.3.4 Diagnosing

Endoscopy and stroboscopy have been used for decades to explore the laryngeal functionality of patients with vocal disorders [12], [13]. While these methods allow surface visualisation of the vocal folds, the movement of laryngeal cartilages and muscles are also important for efficient detection of any vocal disorder. In recent literature, CT technology has been used to detect the pathologies in vocal tract or upper airway [5], [14], [15]. CT provides non-invasive 3D observation of the larynx and modern CT scanners can image the entire larynx in a few seconds without allowing the distortion caused by respiratory motion or articulatory motion.

Currently, the radiologists manually segment the laryngeal structures and measure the parameters in CT images of the larynx. However, modern CT imaging provides large number of images of the larynx and therefore radiologists have to spend lot of time and effort using this manual process. Furthermore, poor reproducibility and accuracy in manual measurement affects correct diagnosis as well.

During this research project, we aim to develop an accurate technique to delineate the vocal folds and related structures from a laryngeal CT image and assist the diagnosis of vocal disorders. To the best our knowledge, there is no known system for delineating vocal folds from CT images in previous literature. We hope this system will provide a fast and accurate way for diagnosing vocal disorders, thus dramatically reducing the workload of radiologists.

1.4 Dataset

Laryngeal CT data used for training, validation and testing were collected from the Movement Disorder Clinic of the Monash Medical Centre. The Ethics Committee of Monash Medical Centre approved all experimental procedures involving human CT images for this project (Application No:11230B). Forty-four idiopathic PD patients and controls (most of the time spouses of the patients) were chosen as the subjects. All the subjects were between 35 to 85 years of age and both males and females were chosen for the study.

Subjects had other disorders along with PD were excluded from our study. Furthermore, subjects had other respiratory disorders like obstructive pulmonary disease, cancers in brain, head or neck and any vocal folds pathologies were also excluded for the data set. Moreover, our controls did not have PD or any other neurological disorders.

PD patients who had PD only for less than 6 years were selected for the patients data set. They were instructed to discontinue their usual PD medication the night before the data acquisition and restart them after the test.

Among the subjects, 16 were known as PD patients' and 17 were controls. The status of the other 11 subjects were not known and used as the uncategorised data-set in the experiment. Among the known patient subjects there were five females and twelve males. Also, among the controls there were seven females and nine males. Similarly, there were five females and six males in the uncategorised data set.

Apart from the recruiting clinicians, other radiologists who performed the CT imaging were blinded to the diagnostic status of the subjects. To analyse the vocal folds movement the laryngeal image data had to be acquired during the phonation. Therefore, all the subjects were asked to execute five short and fast 'i' phonations for 5 s. For every 100 ms, images of the entire larynx were obtained using a 320-slice CT machine.

The images were reconstructed in 3D and the 4D (3D and time) DICOM files obtained were converted in to NifTI-1 format with 512×512 pixels size and 12-bit gray levels with RAS orientation and used in analysis. Here after this data-set will refer as Parkinson-MONASH data-set.

1.4.1   Issues with the Data Set

Some images in the data set were not up to the standards due to various reasons. As we imaged the larynx during vocalisation, some images were blurred and distorted with noise due to the motion of vocal folds and articulating cartilages and muscles.

Posture variations of the patients and the lack of standard starting position for scanning of neck were the other problems in the data-set, and in some cases images of the vocal folds region were not present at all. These images, therefore, were not suitable for analysis.

Moreover, some CT images had quality issues due to low contrast resolution. In these images, distinguishing different tissues was not possible even for a human observer.

These problematic images were removed from the analysis during the experiments.

1.5   Thesis Organisation

Organisation of the rest of this thesis is as follows.

Chapter 2: Literature Review

This chapter presents a summary of background information about the topic. First, existing publications on medical image segmentation methods are summarised. Then a critical analysis of published sources on lower and upper airway segmentation methodologies are described. Finally, a literature review on existing time series analysis in medical imaging is presented.

Chapter 3: Automatic Segmentation of Rima Glottidis using Texture Features

This chapter presents a novel method to segment the rima glottidis from 4D laryngeal CT scans. Texture features calculated on 2D axial CT image slices are used to train a support vector machine (SVM) classifier. Then a robust automatic threshold-based method is proposed for removing the false positives and redundancies and select the optimal detection window. After that, an automatic 2D region growing algorithm was proposed to segment the rima glottidis area. Evaluation of the method is performed using 20 subjects of the Parkinson-MONASH data-set. Classification results and segmentation results are validated with manual gold standard annotations produced by expert radiologists. The

results showed a sensitivity of 87%, precision of 93% at very low false positives per image of 0.0633.

Chapter 4: Estimation of Vocal Folds Plane in 3D Laryngeal CT Images

This chapter presents a reliable and effective method to estimate the vocal folds plane from 3D laryngeal CT scans. An intensity-based method and a rule-based algorithm are used to extract the mid-sagittal plane (MSP) of the CT volume. Then the airway on the MSP is segmented using intensity information. After that, the boundary coordinates profile is used to anterior edge of the airway is extracted and the AC is localised. Next, intensity prior knowledge is incorporated with an unsupervised k-means clustering algorithm to segment the vertebral body column on MSP. Finally, a novel hypothesis-driven method is proposed using anatomical *a priori* to estimate the vocal plane position. Evaluation is performed using 44 subjects of Parkinson-MONASH data-set. Manual markings are provided for AC locations by the expert radiologists. Results shows our method has detected the AC with a 71% sensitivity and 81% precision. However, manual gold standard annotations are not presented for vocal plane positions. Therefore, the results are validated only by visual inspection of the experts and found to be concordant.

Chapter 5: Conclusions

This chapter presents a summary of the contributions this thesis provides to the literature, limitations of the work and suggests potential future research direction for this work.

Chapter 2

Literature Review

This chapter presents a comprehensive background survey about the topic of this research. First, methods used for medical image segmentation in existing literature is described. Then, the main lower and upper airway segmentation methodologies in literature are explained along with a comparison of the results. Finally, a literature review on time series analysis methods in medical imaging are given.

2.1   Medical Image Segmentation

Segmentation is the process of partitioning an image into meaningful smaller pieces, making it easier to analyse the image. In medical diagnosis and treatment planning, accurate image segmentation is a crucial requirement and often requires expert radiologists' knowledge. Increasing use of MRI, CT and other imaging modalities in clinical investigations, treatment planning and monitoring, have given rise to the associated amount of data involved in the studies, hence increasing the requirement of using computer-assisted techniques in medical imaging segmentation process.

Over the last three decades, various general purpose methods have been published to segment the medical images and some of the frequently used methods can be listed as thresholding, edge detection, region growing, level sets, watershed transforms, pixel classification, clustering and deformable models.

Thresholding and Edge Detection

Thresholding is the simplest form of segmentation method which creates binary partitioning of image intensities. Thresholding is effective in segmenting images which have structures with contrasting intensities. However, this technique is more sensitive to noise and intensity inhomogeneities and does not consider spatial characteristics of an image. Due to this reason, this is often used as the initial step to achieve a coarse segmentation before applying more sophisticated segmentation algorithms [16], [17].

Edge-based algorithms also use distinctive intensity or gradient features of the edges of an image to partition the image. Wavelet transforms [18], Canny edge detection [19], and Sobel edge detection [20], are popular edge detection methods used in the literature. Edge detection techniques are more sensitive to noise and partial volume effects in images and therefore often the detected boundaries of the edges may be discontinuous. Morphological operations are used to connect them and eliminate these gaps. Edge-detection methods are hardly used alone and often used as a pre-processing step before the actual segmentation [21], [22].

Region Growing

Region growing is a simple algorithm proposed by Adams and Bischof [23], which starts each region with a single seed pixel and then decides whether the neighbouring pixels should be added to the region depending on the homogeneity criterion. Different varieties of region growing techniques can be found in the literature [24], [25]. The decision-making criterion can be chosen to suit the application type.

Despite the simplicity and efficiency of the algorithm, region growing always suffers from leakage problems or under segmentation issues due to noise, artefacts and partial volume effects in medical images. Furthermore, robust and efficient identification of an initial seed is not always an easy task. Therefore, region growing is always implemented as the initial step combining with other segmentation techniques to achieve the best re-

sults [25], [26], [27].

Level Sets

Level set methods use partial differential equations depending on different features of image and define positive values inside the segmentation and negative values outside. Therefore, the segmentation boundary is defined as the zero level set. This method has the capabilities to handle complex topological and geometric changes efficiently in 3D and thus widely used in detecting tumours in MRI images of brain structures [28].

Watershed Transforms

Watershed algorithms treat an image as a topographic surface and the watershed lines are considered as the pixels with the local maximum gradient magnitudes. Pixels surrounded by same watershed lines are considered as a region. The algorithm searches the pixels with the lowest gray-level and propagates the water in that direction. Then the pixels that classified as local minima are grouped into the same region. This algorithm uses diverse information in an image and therefore provides an effective complete segmentation.

However, when the signal-to-noise ratio is low this produces over-segmented results. Therefore, hybrid segmentation techniques incorporate other methods with watershed transforms to achieve better results [29]. Grau et al. [30] and Hamarneh et al. [31] used priori knowledge in watershed method and improved the performance of the algorithm.

Pixel Classification

Classification is a supervised method of segmentation, which partitions a feature space derived from an image using a set of training data, which have known features and class labels. Features can be any known functions of an image which provide the best separation between the region of interest and the background. Pixels with unknown labels are grouped in to the classes according to the maximum probability of that pixel belonging

to a certain class.

The kNN classifier (k-nearest neighbour classifier) is the simplest classifier and it groups the pixels in to the same class as the training data based on the similarity of the intensity [27]. One disadvantage of this method is that the requirement of manual intervention to train the data. Manual labelling of data for each class is needed and can be time consuming.

Clustering

Clustering method has the same functionality as classifiers. However, it is an unsupervised method and does not need any training data. The algorithm trains itself using available data while doing the segmentation at the same time. However, this method requires some initial segmentation [32].

Deformable Models

Deformable models are model-based segmentation methods that are widely applied in medical image segmentation. These are closed parametric curves/surfaces used to delineate the region boundaries by deforming the under application of internal or external forces. Initially, the closed curve/surface is placed inside the region of interest and the internal forces computed within the curve/surface are applied, allowing the boundary to relax iteratively.

Deformable models are often used in reconstructing cerebral cortex from MRI images, cardiac image segmentation and bone CT image segmentation. Active contours is the first deformable model applied to medical image segmentation [33]. The main disadvantage of deformable models is that it needs some manual input to locate the initial contour and select the initial parameters.

A summary of some medical image segmentation techniques is shown in Table 2.1.

Table 2.1: Comparison of existing medical image segmentation literature

Study	Year	Methods	Data	Type	Comments
Adams et al. [23]	1994	Region growing	2D X-ray	Semi Auto	Application in lung segmentation; sensitivity to the initial seed
Gordon et al. [16]	1996	Thresholding, Region growing	2D/3D CT	Semi Auto	Segmented the trabecular bone
Singleton et al. [17]	1997	Thresholding, Recursive region growing	2D/3D MRI	Auto	Segmented cardiac MRI images with 90.3% efficiency
Hojjatole-slami et al. [24]	1998	Region growing based on gradient information	2D/3D MRI/CT	Semi Auto	Insensitivity to noise; independent from the initial seed
Grau et al. [30]	2004	Improved watershed transforms based on priori knowledge	3D MRI	Semi Auto	For segmentation of knee cartilage and grey/white matter in brain; similar or superior performance over manual segmentation
Ng et al. [32]	2006	k-means clustering, Improved watershed transforms	2D MRI	Auto	Segmentation map with 92% fewer partitions than conventional watershed
Andreao et al. [21]	2007	Edge detection, Wavelet transforms, Hidden Markov Models	2D ECG	Semi Auto	Beat segmentation and premature ventricular beat detection results were comparable to the previous literature
Qin et al. [22]	2007	Edge detection, Level sets method	2D/3D Virtual human images	Semi Auto	Improved speed; improved precision segmentation while keeping the edges smooth

Table 2.1 – *Continued from previous page*

Study	Year	Methods	Data	Type	Comments
Rusko et al. [26]	2009	Contrast enhancement, Region growing, Knowledge-based methods	2D/3D CT	Auto	For liver segmentation; over 90% cases reported comparable results with manual standard
Van et al. [27]	2009	Region growing, Pixel classification	2D/3D CT	Auto	For segmentation of Lungs, Lobes, Airways, Fissures and segments; accuracy of 77% (compared to inter-observer - 74% and intra-observer - 80%)
Aloui et al. [28]	2009	Level sets method, Meshes simplification	3D MRI	Semi Auto	For brain tumour delineation; satisfactory results
Hamarneh et al. [31]	2009	Watershed transforms based on prior shape/appearance knowledge, k-means clustering	2D/3D MRI	Auto	Slightly different results with manual expert segmentation
Dongxu et al. [25]	2013	Thresholding, Seedless region growing, Rule-based methods, Morphological Reconstruction	2D/3D MRI	Auto	For mandibular body segmentation; improvement in accuracy over conventional region growing and level sets

2.2   Airway Segmentation

Accurate segmentation of human trachea is essential in analysing airway-related diseases and calculating and visualising laryngeal parameters. Although manual segmentation of

airway gives the most accurate measurements, tracing the boundaries slice by slice and reconstructing them into a 3D volume requires a lot of time and specialist intervention. Automatic segmentation schemes, however, drastically reduce the time [34]. A number of studies have been published in the literature to segment upper airway and lower airway in different imaging technologies, especially with MRI and CT. Many of them focus on segmenting the lower airway trees.

Despite the fact that MRI provides images with better tissue contrast without any radiation damage to the patient, using MRI to acquire laryngeal/thorax images will introduce some amount of blur in the images due to unavoidable movements like respiration or articulations during its long acquisition time. Particularly, in upper airway MRI images the calcified structures which contain some amount of hydrogen are difficult to distinguish with the air lumen. In contrast, the modern CT scanners acquire laryngeal image volumes during a single breath-hold time with a good spatial resolution. Due to this reason, most of the existing literature have used CT images for airway segmentation. However, this is at the expense of a relatively high radiation dose to the patient.

Airway is a cylindrical structure with a decreasing radius from the nasal cavity or lips to the lungs. Generally, the upper part of the tract from the vocal folds (glottis) is known as the upper airway or the vocal tract and the part from the glottis to the lungs is known as the lower airway.

2.2.1 Lower Airway Segmentation

According to the characteristics of the methodology used, lower airway segmentation schemes can be divided into six categories: (i) region growing/wave propagation; (ii) knowledge-based/rule-based segmentation; (iii) template matching methods; (iv) machine learning methods; (v) morphological methods and (vi) geometric shape analysis methods.

Region Growing/Wave Propagation

Due to the high contrast between the air and the tube wall, the 3D region growing is the most often used method to segment the airway region from CT images. However, due to partial volume effects and noise artefacts, segmentation solely on region growing method will often result in early termination of segmentation process or leaking in to the lung tissue. Although some methods like front propagation [35] have been developed to prevent the leakage problems, complete prevention is impossible. A threshold and an initial seed location should be provided in order to start the region growing process.

Mori et al. [36] in 1996, introduced a method which gradually increased the threshold applied until the total number of voxels showed a sudden increase which indicated a leakage in to the lung tissue. However, this method assumed that the airway wall was not broken at any place, which practically is not the case. Due to noise or pathologies, there are often smaller regions existing on the wall, where the wall is invisible and the air lumen and the surrounding tissue has the similar intensity. Attempts to stop these local leakages and to allow the region growing process to continue to the other regions have been carried out by several authors [35], [37], [38], [39].

Schlathoelter et al. [35] was able to segment all major airway branches and most of the minor branches using front propagation approach. In this method, the diameter of the airway tube was measured and a leak was identified when the diameter showed a sudden increase. Then in 2003, Kitasaka et al. [37] improved Mori's method [36] by preprocessing the airway tree by applying a sharpening operation and then using anatomical knowledge to reduce leakages and achieved better segmentation. Singh et al. [40] used information gain at each voxel as the decision criteria for adaptive region growing. Although, this method was able to produce acceptable results, it was not free from leakages. Tschirren et al. [38] used smaller adaptive regions of interest that followed the airway tree as they were segmented. This method detected possible leakages early and avoided them, thus reducing large-scale leakages that had occurred in previous methods. However, this method is sensitive to the motion artefacts, such as heart beat or respiration and also to the pathologies which obstruct the airway.

Deciding an optimal threshold which prevents the leakages and allows the maximum segmentation is always a difficult task in region growing. Also, region growing depends on the effective initial seed selection which most of these methods use to search a circular, air-filled area in 2D image slices. Whereas, it is not always guaranteed that the areas satisfying these specifications represent the trachea. Due to these limitations, region growing is always implemented along with some other techniques to achieve better segmentation. In 2008, Nakamura et al. [41] proposed a similar method to [36] by adaptively changing the threshold value by using the previously segmented voxels. This method incorporated a Hounsfield values-based correction algorithm to correct the diameters of the airway branches and ensure the maximum segmentation.

Knowledge-based/Rule-based Segmentation

Knowledge or rule base methods incorporate the known anatomical or functional properties such as shape, decreasing diameter, angle and intensity of the airway to the segmentation methods to achieve better results. Sonka et al. [42] applied the anatomical knowledge of presence of an airway wall around the airway lumen and the pair-wise appearance of airway and blood vessels in 3D region growing method. As vessels always appear brighter than lung tissue in a CT image and can distinguish more easily than airways, this method gave a sensitivity of 69%-84% compared to the 48% of the conventional 3D region growing method. However, as the existence of adjacent airways to the vessels is not always guaranteed this method resulted in a large number of false positive (FP) detections.

In 1998, Park et al. [43] improved [42] by incorporating a fuzzy logic decision-making system that evaluated the low brightness, adjacency to the vessels and the degree of wall existence to determine the confidence level of being a region of an airway. This method did not change the sensitivity of the previous method, but reduced the number of FPs. Despite this improvement, this method required some manual interaction to design fuzzy logic functions and also assumed a connected 3D CT data volume, which may not always be available.

In [37], we can see another approach to segment the airway using the anatomical knowledge of having a decreasing diameter. Tschirren et al. [38] used the knowledge that no isolated regions exist in the airway (fuzzy connectivity method) for segmentation and avoided most of the large-scale leakages in the conventional region growing approach. Another rule-based approach is described in [44], which used the knowledge of degree of existence of the air wall, gradient index, mean gray-level and the ratio of holes in a fuzzy logic decision-making system to detect the airways. Although this method outperformed the conventional region growing segmentation method, the method still had some minor leaks. In 2013, [45] adapted a rule-based hybrid region growing algorithm to segment the airways. Vesselness and the adjacency of two voxels in image space were calculated and used to detect the airway voxels. This method was fully automatic and detected most of the airway branches even under pathologies, noise and artefacts efficiently with less FPs.

In practice, it is difficult to define all the rules which give the appropriate level of confidence for being an airway. Therefore, depending solely on these rules or knowledge is not sufficient for a robust airway detection system. It is always essential to incorporate other techniques to remove the FP detections.

Template Matching Methods

Airway template matching methods search for airway regions that match with predefined shapes. Normally for airways, 2D circular shapes with range of sizes and intensities are used as templates and the algorithm looks for these shapes in each slice.

Bartz et al. [46] combined 2D template matching used varying size circular templates with 2D region growing to identify smaller airway branches. However, the variability of airway sizes from patient to patient makes it difficult to stick to a few sizes of templates, thus increasing the time needed for the segmentation process. Furthermore, airway appears in circular shapes in 2D slices only when the travelling direction of air is perpendicular to the image plane. Therefore, the airway in a different 2D slices can be elliptical shapes with various sizes, which makes the template matching more complex.

In 2005, Tschirren et al. [38] proposed to use smaller templates of cylindrical shape, to overcome this problem. This prevented most of the large-scale leakages. Similarly, Bulow et al. [47], Grahan et al. [48] and Van et al. [39] also used small cylindrical tube shapes matching in their work to achieve the segmentation. However, due to the attachment of airways to the surrounding structures, these template matching methods have a high probability of giving inaccurate detections.

Machine Learning Methods

Machine learning methods represent the data and the priori knowledge in the best possible way and thus knowing the possible patterns of true airways it captures the probability distribution of the airway features in the image.

Ochs et al. [49] used a voxel classifier method which classified various lung structures including the airways from CT images using a single small feature set. They used voxel attenuation and Eigen values of the Hessian matrix as the features and the training of the classifiers was done by a set of experts by manual labelling the lung structures. The method was found to perform better on thin image slices, where the partial volume effect was minimal. Airways were misclassified with edge parenchyma due to the appearance of similar airway-like structure near the edge of the lung tissue. It was found out that if the classifier was trained to a specific image structure rather than multiple structures, better performance could be achieved. However, the aim of this study was to introduce a generic method which segments multiple structures that can be used as an input to a more advanced segmentation method.

Lo et al. [50] introduced an approach which used a kNN classifier and a large appearance-based feature set, which produced better differentiating of airway voxel and non-airway voxels at different scales with less leakages. Due to the difficulty of finding a high quality manually segmented training data-set Lo used the output of an interactive region growing method to train the classifier. Although the proposed method was able to capture some missing airway branched in the imperfect training data-set, the number of FPs were significantly large. Also, this method did not perform better in segmenting the airways

which were larger than the average size. In 2009, Lo et al. [51] improved their previous method by introducing an optimal path search method and reported 36% increase in the detected airway tree length with a slight increase in the false positives rate (FPR). In 2010, Lo et al. [52] again improved [50] by introducing a vessel orientation similarity measure, which showed how similar the airway is to the neighbouring vessel, to the classifier-based appearance model and reported an average 7% increase in branch extraction.

Morphological methods

Morphological methods identify parts that have a likelihood of being an airway by analysing the shape and size in each slice and the spatial relationship between the slices. Therefore, generally these methods search for circular regions in 2D slices and then reconstruct the 3D airway using various morphological operations.

In 1996, Pisupathi [53] used mathematical morphology to extract airways, pulmonary artery, pulmonary veins and lungs. Using various size structuring elements, he segmented each 2D slice separately and then reconstructed the 3D tree. However, this method was used in canine lungs and never tested on images of human lungs.

In 1997, Preteux et al. [54] combined watershed transforms and connection cost-based marking with morphological filtering to extract the airways of sheep. They focused mainly on segmenting the 2D slices rather than reconstructing the 3D airway tree. Later, Fetita et al. [55] improved this method by introducing 3D airway reconstruction using morphological reconstruction and tested the method in human lungs.

In 2003, Aykac et al. [56] proposed a method which identified the local minima in 2D CT image slices, using a range of gray-scale morphological structuring elements and then reconstructed the 3D airway tree using closed space dilation. The method reported 73% detection of airway branches compared to the manual gold standard. However, the method was sensitive to the structuring elements used and did not define the airway borders accurately. The objective of the study was to reconstruct the airway and not to define the luminal borders.

Kiraly et al. [57] showed that integrating the adaptive region growing with the mathematical morphology method reduces the time required for the morphological method alone and gives better edge localisation and effective segmentation.

In 2004, Fetita et al. [58] developed a novel method in contrast to the conventional morphological method, by introducing two morphological operators, selective marking and depth-constrained connection cost, to use in the 3D energy-based growth process and showed better performance even under severe obstructive airway pathologies. Fabijanska [59] used morphological gradient constrain in 3D-seeded region growing and achieved satisfactory results in airway segmentation. However, the accuracy of this method was dependent on the level of pathology presented.

Reconstruction of 3D data by morphological methods greatly depends on the availability of continuous data in 3D space. Otherwise, the 3D reconstruction of the extracted 2D regions will produce discontinuities.

Geometric Shape Analysis Methods

As airways are tube-like structures, several previous studies have used geometric shape-based methods which use the Eigen value analysis of the Hessian matrix [60], [61], [62] to detect tubular structures in the image. However, use of the second order derivative information in the Hessian matrix makes the result more sensitive to the noise. Furthermore, before applying Eigen value analysis, these methods apply Gaussian filtering to the original image to remove the noise. This introduces some amount of blurring to the image, ultimately leading to miss the smaller airways from the segmentation. Moreover, using the gradient information in the linear-scale space produces diffusion of neighbouring structures in to one another, thus adjacent multiple tubular structures are recognised as a single object.

In 2008, Bauer et al. [63] proposed using the gradient vector flow instead of the gradient vectors which avoided the interferences from the adjacent structures. They combined

the Frangi's vesselness measure [64] with gradient vector flow and detected the centre-line of the tubular structure which was not dependent on the size or the contrast of the tube. In 2010, Graham et al. [65] used a shape analysis method, that searched for small elliptical structures in axial, sagittal and coronal slices of the CT volume and reconstructed the 3D airway by joining them. However, this method was sensitive to slice thickness and patient motion.

Jiantao et al. [66] introduced a new method which used principle curvature and principle directions to identify the airway branches from the surrounding tissues in geometric space. This method was not sensitive to noise or artefacts and virtually free from leakages. Also, it was able to detect the isolated branches due to pathological or other obstructions.

As we can see any of the known segmentation techniques have not been able to achieve a perfect segmentation of the airway tree without leakages or FPs. Almost all the work in the literature has combined two or more methods to achieve better results.

A comparison of known previous methods is shown in Table 2.2. However, as the data-sets, testing criteria and the reference standards were different for each case, comparing the performance of the different methods could not be achieved simply in this way.

Table 2.2: Comparison of existing lower airway segmentation literature

Study	Year	Methods	Data	Type	Comments
Mori et al. [36]	1996	Adaptive region growing based on optimised threshold	3D CT	Auto	Failed to extract very thin bronchi
Sonka et al. [42]	1996	Region growing, Rule-based methods	2D/3D CT	Auto	True positives 68.5%; false positives 11.5%

Table 2.2 – *Continued from previous page*

Study	Year	Methods	Data	Type	Comments
Pisupathi et al. [53]	1996	Mathematical morphology, Knowledge-based methods, Region growing	2D/3D CT	Auto	25% increase in detected voxels compared to thresholding; did not miss any major airway branches
Preteux et al. [54]	1997	Thresholding, Mathematical morphology, Watershed transforms	2D/3D High reso- lution CT	Semi Auto	Accuracy up to 90%
Park et al. [43]	1998	Region growing, Fuzzy-logic rule-based methods	2D/3D CT	Auto	True positives 69.1%; false positives 8.8%
Sato et al. [60]	1998	Mathematical models	3D MRI/CT	Semi Auto	Bronchi can be more reliably classified
Krissian et al. [61]	2000	Mathematical models	3D MRI/2D X-ray	Semi Auto	More precious contour detection was obtained
Fetita et al. [55]	2000	Mathematical morphology, Energy-based modelling	2D/3D CT	Auto	Robust with respect to anatomical variabilities; improved accuracy
Kiraly et al. [57]	2002	Region growing, Morphological dilation	2D/3D CT	Auto	More True positives and False positives; up to 12th order bronchi; 2-25min for total segmentation
Kitasaka et al. [37]	2003	Edge sharpening, VOI extraction, Region growing	3D CT	Semi Auto	Reduced large scale leakages; 4th order bronchi - 82%; 5th order bronchi - 49%; 6th order bronchi - 20%

Table 2.2 – *Continued from previous page*

Study	Year	Methods	Data	Type	Comments
Bartz et al. [46]	2003	Region growing, 2D wave propagation, 2D template matching	2D/3D Multi-slice CT	Semi Auto	Up to 5th order bronchi - more than 85%; Up to 6th order bronchi - more than 58%; After 7th order bronchi - below 30%; Total segmentation time 20-100s
Li et al. [62]	2003	Filter enhancement, Mathematical models	2D/3D CT	Semi Auto	Reduced false positives
Aykac et al. [56]	2003	Morphological reconstruction (closed space dilation, gray-scale closing)	2D/3D CT	Auto	85%-89% branch matches with the manual results; sensitivity 73%; Total 364 branches
Bulow et al. [47]	2004	Template matching, Wave front propagation	2D/3D Multi-detector CT	Auto	Application to the coronary artery detection
Singh et al. [40]	2004	Gain-based region growing	3D CT	Auto	Segmentations were closely correlated with manual segmentation; not free from leakages
Fetita et al. [58]	2004	Mathematical morphology, Energy-based reconstruction	3D Multi-slice CT	Auto	Sensitivity 91%; robust with pathologies
Tschirren et al. [38]	2005	Fuzzy connectivity, Region growing	3D CT (Low dose)	Auto	For total lung 3 min$\pm$30s; 100% segmentation up to 3rd order; sensitive to motion artefacts; 27.0$\pm$4.4 segments per tree

Table 2.2 – *Continued from previous page*

Study	Year	Methods	Data	Type	Comments
Ochs et al. [49]	2007	Voxel classification	3D CT	Auto	Single feature set; 92% of voxels of larger airways; 86% of voxels of smaller airways were correctly identifies
Van et al. [39]	2008	Multi-threshold adaptive region growing, Wavefront propagation, Rule-based	3D CT	Auto	Segmented up to 10th order without leakages; 10 s per case; 90% sensitivity
Lo et al. [50]	2008	Voxel classification (appearance model-based kNN classifier), Region growing	3D CT	Auto	No leakages; higher true positives rate (>90%); false positives rate 1.35%; reduced performance when airways were larger than average
Nakamura et al. [41]	2008	Region growing with HU value-based correction, Morphometric analysis	3D Multi-detector CT	Semi Auto	Up to 14th order bronchi
Graham et al. [48]	2008	Graph optimisation technique	3D Multi-detector CT	Auto	Low false positives rate; high sensitivity to peripheral airways
Bauer et al. [63]	2008	Geometric shape analysis	3D MRI/CT	Semi Auto	High robustness to noise; detects vessels independent of size or contrast
Tan et al. [44]	2009	Fuzzy logic rule-based segmentation	2D/3D CT	Auto	More branches and less leaks; up tp 6th order bronchi

Table 2.2 – *Continued from previous page*

Study	Year	Methods	Data	Type	Comments
Lo et al. [51]	2009	Voxel classification (appearance model-based kNN classifier), Optimal path technique	3D CT	Auto	Slightly increased False Positives rate (3.79%); Increase of 36% in detected tree length
Fabijanska et al. [59]	2009	Region growing constrained by morphological gradient	3D Multi-detector CT	Auto	Average branch detection 37%; average tree length detected 36%; average false positive rate 1%; average leakage count 2.3; depends on the level of pathology; up to 9th order bronchi; Total time per segmentation – 10 min
Graham et al. [65]	2010	Region growing, Graph optimisation	2D/3D Multi-detector CT	Semi Auto	> 7th generation; time for segmentation –3 min
Lo et al. [52]	2010	Voxel classification (based on vessel orientation similarity), Region growing	3D CT	Auto	50% less leakages; 20% longer tree detections
Jiantao et al. [66]	2011	Thresholding, Geometric shape analysis	3D CT	Auto	Insensitive to noise and artefacts; no leakages; average total airways length of the segmented airway trees per examination was 261.8 cm; 30 min for complete airway

Table 2.2 – *Continued from previous page*

Study	Year	Methods	Data	Type	Comments
Ziyue et al. [45]	2013	Fuzzy connectivity, Region growing	2D/3D CT	Auto	Not sensitive to noise and artefacts; low false positives (0.19%); detected tree length – 48.6%

2.2.2 Upper Airway Segmentation

While there is much work published to segment the lower airway, comparatively little literature has been published for lower airway segmentation. Although the upper airway segmentation is a challenging task with the air-filled areas inside the head and due to the dynamic nature of the articulators, it is not as complex as the lower airway due to the absence of the minor branches.

In 1999, Fried et al. [67] used an adaptive segmentation of the upper airway using Markov random field model in both CT and MRI images. However, due to the significant time consumed by the algorithm, this was only conducted in 2D domain. Then, in 2001 Dongqing et al. [68] proposed a 3D segmentation approach to extract the vocal tract, by generating a series of feature vectors for each voxel and applying a self-adaptive vector quantisation algorithm to those feature vectors to classify the vocal tract voxels. The method was applied to both CT and MRI images and received satisfactory results.

In 2003, Liu et al. [69] proposed a method to diagnose upper airway diseases of children by segmenting the upper air tract from MRI images using fuzzy connectedness between the voxels. The accuracy of the segmentation was 97% and the result was used to study the size, shape, volume and area of the airway.

In 2005, Ogawa et al. [14] used a simple gray-scale thresholding to delineate the upper airway from the surrounding tissue in cone beam CT images. This method needed

an intensive manual interaction to select the region of interest from the image slices. Shi et al. [70] improved [14] by changing the threshold value and reducing the user interaction by introducing a uniform quadrangle to select the region of interest. However, this method was not validated in the patients with upper airway pathologies.

Cheng et al. [15] tracked the upper airway by combining edge detection and gradient vector flow snakes technique and using the contour of the airway in the previously extracted slice as *a priori*. In 2009, Bresch et al. [71] used the spatial frequency domain information to extract the upper airways from MRI images. This method fitted the image data to an anatomically informed unsupervised object model.

Recently, Raeesy et al. [72] used oriented active shape models based on defined shape models to delineate the vocal tract from the surrounding tissues that have similar properties. Therefore, this method was able to identify the airway among other air-filled regions, especially which are in the patient's head.

A comparison of the existing upper airway segmentation techniques is shown in Table 2.3.

Table 2.3: Comparison of existing upper airway segmentation literature

Study	Year	Methods	Data	Type	Comments
Fried et al. [67]	1999	Markov Random Fields, manual segmentation	2D/3D CT/MRI	Semi Auto	Application in virtual endoscopy; Provided good representation of laryngeal neck anatomy; Further validation needed
Dongqing et al. [68]	2001	Self-adaptive vector quantisation algorithm for feature vectors	3D CT/MRI	Auto	Segmentation of vocal tract for virtual laryngoscopy; insensitive to mild motion
Liu et al. [69]	2003	Fuzzy connectedness	3D MRI	Auto	Accuracy 97%; mean time per study - 4 mins
Ogawa et al. [14]	2005	Gray-scale thresholding	3D Cone beam CT	Semi Auto	Application in diagnosing of obstructive Sleep Apnoea
Shi et al. [70]	2006	Gray-scale thresholding with uniform quadrangle to select region of interest	3D Cone beam CT	Auto	Very small difference in results compared to manual segmentation; not validated against upper airway pathologies
Cheng et al. [15]	2007	Edge detection, Gradient vector flow snakes	2D/3D Cone beam CT	Auto	For the given data-set all the contours detected were acceptable; may be sensitive to the breathing motion
Bresch et al. [71]	2009	Anatomicaly informed unsupervised model-based method	2D MRI	Auto	Outperformed the manual segmentation; effective in detecting air/tissue boundary in vocal tract
Raeesy et al. [72]	2013	Oriented active shape models	2D MRI	Auto	Average precision of 78.5%

2.3   Time Series Analysis In Medical Imaging

Time series analysis is used in medical imaging domain frequently to predict the onset of a disease, analyse the progressive pattern of a disease or the treatment effect of a disease. In medical imaging domain, time series data means 4D image data which consists of multiple image volumes at different times.

In recent years, number of approaches have been developed to quantitatively analyse the motion patterns in anatomical structures and to detect the pathological structures. In 2005, Perperidis et al. [73] developed a 4D cardiac statistical atlas which represented the cardiac anatomy during the cardiac motion, using the distribution spaces of the cardiac shape due to inter- and intra-subject variability. Calculated modes of variation in each distribution using principle component analyses were used in the K-weighted NN classifier to differentiate between cardiomyopathy patients and normal subjects.

In 2010, Honghai et al. [74] introduced a method to analyse 4D cardiac MRI images using motion and shape features. This method computed normalised volume-time curves and volume changing rate (dV/dt) curves for the segmented left and right ventricles and extracted the model indices using principle component analysis on the curves. Then, those indices were fed in to a classifier to differentiate the healthy and pathological hearts. The abnormal motion pattern in the pathological heart was accurately captured and reported 90%-100% tracking sensitivity. However, these methods need a significant amount of training database to handle the large-scale variations in the spatial domain.

In [75] a novel method has been proposed to detect and quantify the pathologies in 4D data of human wrist bone. In this method, the relative position of points on adjacent bone surfaces was used as the descriptor to analyse the wrist movement pattern. As this motion followed the Gaussian distribution, pathological wrists were detected based on the difference in distances between the points.

In 2014, Rahman et al. [76] analysed a time series of high speed digital images of human vocal folds to detect pathological conditions. They computed Glottal Area Waveform (GAW), which is the plot between the area of the vocal folds opening and time,

for different subjects and used dynamic time warping classifier to match the each subject's GAW with a reference pattern and classify the abnormal vocal folds. The accuracy of this system depended on the reference templates and the quality of the GAW acquired.

Existing literature in the time series analysis area is listed in Table 2.4.

Table 2.4: Comparison of existing time series analysis literature

Study	Year	Data	Methods	Comments
Perperidis et al. [73]	2005	4D CT	Distribution space of cardiac shape, k-weighted NN classifier	Differentiated cardiac myopathy patients and normal subjects
Honghai et al. [74]	2010	4D MRI	Normalise volume-time curves, Classifier	Tracking sensitivity of 90%-100% in detecting abnormal pathological heart
Van et al. [75]	2012	3D rotational X-ray in different poses	Based on relative position of points on bone surfaces	Detected pathological wrist bones
Rahman et al. [76]	2014	Time series of high speed digital images	Glottal area waveform, Dynamic time warping classifier	Differentiated abnormal vocal folds

Chapter 3

Automatic Segmentation of Rima Glottidis using Texture Features

Rima glottidis is the narrowest part of the human laryngeal cavity, which exists between the true vocal folds. In this chapter, we propose a novel robust machine learning model to segment the rima glottidis and generate the time series curves of rima glottidis area from 4D laryngeal CT scans. This work is based on the following publication.

S. Hewavitharanage, J. D. Thyagarajan, K. Lau and M. Palaniswami, "Automatic segmentation of the rima glottidis in 4D laryngeal CT scans in Parkinson's disease," 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, 2015, pp. 739-742.

3.1 Introduction

Due to various causes like neurological disorders, structural abnormalities or aging people may suffer from different voice-related disorders such as hypophonia(reduced speech volume or intensity), dysphonia (poor voice quality), hypokinetic articulation (reduced range of articulatory movements), dysprosodia (reduced pitch), rush (tendency of speech articulation to festinate), hesitant and disfluency in speech [8], [9], [10]. These voice disorders occur when one or both vocal folds cannot move correctly due to reduced or absent nerve input to the laryngeal muscles or damaged vocal muscles. Vocal

fold paralysis can be unilateral where a single vocal fold cannot open or close correctly or bilateral where the both vocal folds cannot move. Unilateral vocal fold paralysis is the most common disorder while the bilateral vocal folds paralysis can be a life-threatening condition as it affects swallowing and breathing.

In either case, the phonation is the most affected part of the speech. The phonatory process occurs when air is expelled from the lungs through the glottis, creating a pressure drop across the larynx. Normally, the vocal folds remain closed until this laryngeal pressure is sufficient enough to draw them apart. When they are opened, air escapes through the glottis reducing the pressure enough for the muscle tension to pull the folds back together again. This abduction and adduction of vocal folds are controlled by the muscles connected to thyroid cartilage, cricothyroid cartilages and arytenoid cartilages. When these muscles and vocal folds are impaired due to some abnormality, it will result in an incomplete glottal closure and hence the gap between the vocal folds (rima glottidis) is also affected [7].

Previous endoscopic and stroboscopic studies on a group of PD patients, showed impaired glottic competence and vocal folds bowing in PD patients which causes increased glottic area during vocalisation [77], [78], [79]. Furthermore, studies done by Perju et al. [80] in 2014 using the CT images of larynx of PD patients also confirm this hypothesis. Therefore, this feature can be used to do a quantitative analysis to detect PD patients at early stage [11]. During this project, we developed a technique to delineate the rima glottidis from 3D laryngeal CT scans of human subjects to assist diagnosis of vocal disorders.

Current techniques of measuring the laryngeal parameters involve manual tracing of the object boundaries of the axial CT images taken during the phonatory process of an individual with vocal disorder, and radiologists measuring the distances. Modern CT imaging technology provides 3D images of the entire larynx and therefore a large number of data-sets are usually exist for analysis. Due to this reason, the manual segmentation process takes a lot of time, effort and specialised labour but is often poor in reproducibility. Furthermore, due to inter- and intra-observer variability in results the accuracy in

outcome is always questionable.

In our work, we used a texture feature-based classifier to segment the vocal folds region from the 3D laryngeal CT image. A region growing algorithm was applied to the classified region, which resulted in accurate delineation of the rima glottidis area. The graph developed between the segmented area and the time can be used to predict the presence of voice disorders. The automatically calculated area in different temporal frames from our system are compared with the manually segmented areas and appear to be concordant.

To the best of our knowledge, there is no known system for delineating vocal folds from CT images in previous literature. We hope this system will provide a fast and accurate method for early detecting voice-related disorders, thus dramatically reducing the work load of radiologists.

3.2   Texture Features

Texture is a measure of intensity variation, appearance and arrangement of elements (i.e. pixels) in an image. It provides useful information about the smoothness, coarseness, fineness and homogeneity and helps to differentiate objects which cannot be visually separated. Texture analysis can be divided into four categories, structural methods, model-based methods, transform methods and statistical methods [81].

Structural methods define the texture by well-defined primitives and their spatial arrangement [82] and try to describe the image as a geometric representation. Structural analysis often uses mathematical morphology to analyse the images. While this methodology provides a good symbolic description of the image, this is more difficult when using real images rather than artificial images. Examples of the primitive shapes used for structural analysis are image edges, different shapes and polygons. However, by using more complex shapes, more general spatial relationships can be calculated. Spatial relationships are computed using the periodicity and the regularity of the texture elements.

Model-based methods define the structure using a mathematical model and estimated model parameters are used to analyse the image texture [83], [84]. These methods are relatively mathematically complex to compute. Most popular models used for texture analysis are fractal models, Markov-random fields and autoregressive models. Fractal models are suitable to analyse the texture of naturally occurring scenes and in medical images [85]. Fractal is a complex texture pattern repeatedly occurs at various scales irrespective of the scale. As most of the texture shows some degree of similarity at various scales this provides a good representation of the geometric complexity of the image texture. Autoregressive models and Markov random fields provide a more complete analysis for the image texture. However, estimating the model parameters in these models is extremely complex. Maximum likelihood approach or least error method is commonly used to estimate the parameters in autoregressive models [86]. Most of the time the local texture of an image is dependent on the intensity of the neighbourhood pixels. Using this feature, Markov random fields defines the image texture using the conditional probabilities of the neighbourhood [87].

Transform methods analyse the image texture by transforming it to another space like frequency or scale. These methods normally use Fourier transforms, Wavelet transforms or Gabor transforms to solve the problem [88], [89]. Wavelet transforms are more commonly used due to the flexibility in usage in different problems [90]. These spectral methods normally use a set of filters to transfer the image texture to the frequency domain. The filter can be chosen to enhance some features and suppress unwanted features.For an example, Laplacian filters can be used for edge detection, Gaussian filters can be used for blob detection and Gabor filters perform well in detecting directional texture. However, choosing the filter type is usually heuristic and sometimes the feature data may consist of some unwanted information as well.

Statistical approaches represent the image texture by statistical properties of the gray-level relationships of the image pixels and provides higher discriminative power compared to the other methods [91], [92]. In the statistical method, a few types of features, first order, second order and higher order statistics, can be derived from the images.

3.2.1 Statistical Features

First order statistics

The simplest form of features is the first order features which depend on the gray levels of the individual pixels, such as mean and variance. This uses the histogram of intensity levels, which shows the total number of pixels in each intensity level of the image, to calculate the image properties. The histogram contains the simple statistical information about the image. Let's say $I(x, y)$ is the intensity image function and the image size is $N \times M$ where $x = 0, 1, ..., N - 1$ and $y = 0, 1, ..., M - 1$ and the intensity histogram is given by $H(i)$. When the number of gray levels in the image is N_G the intensity level at each pixel of image can take $i = 0, 1, ..., N_G - 1$, $H(i)$ gives the total number of pixels have the ith intensity level. We can calculate the probability density function (PDF) of intensity levels, $p(i)$, by dividing the histogram by the total number of pixels in the image.

$$p(i) = H(i)/(N \times M) \tag{3.1}$$

Therefore, the first order statistics of the image can be computed as follows.

$$Mean = \sum_{i=0}^{N_G-1} i p(i) \tag{3.2}$$

$$Variance = \sum_{i=0}^{N_G-1} (i - Mean)^2 p(i) \tag{3.3}$$

$$Average\ energy = \sum_{i=0}^{N_G-1} p(i)^2 \tag{3.4}$$

$$Entropy = \sum_{i=0}^{N_G-1} -p(i) log_2 p(i) \tag{3.5}$$

Mean gives the average intensity level of the image being considered and the variance is the spread of intensity level from the mean. Entropy measures the randomness or uniformness of the image. In actual images, these properties do not give much information about the image texture. They rather carry information about the conditions during the image acquisition, such as lighting. Furthermore, first order statistics consider only a single pixel and do not take the relationships between the pixels into account. Therefore

these measures are more sensitive to the changes in the image.

Second order statistics

To avoid the problems with first order statistics and to achieve a better classification of image texture second order statistics are used. Second order statistics consider the relationship between two pixels of the images to calculate co-occurrence matrices or gray-level difference matrices (GLDM). Feature descriptors such as energy and entropy are derived based on this gray-level co-occurrence matrix (GLCM), GLDM or autocorrelation function.

Gray-level co-occurrence matrix: GLCM is first introduced by Haralick et al. in 1973 [91] as spatial gray-level dependency matrix (SGLDM). This computes second order texture features and has proven to achieve similar level of classification performance as human vision. For an image with size $N \times M$ where $x = 0, 1, ..., N - 1$ and $y = 0, 1, ..., M - 1$ and gray levels N_G, GLCM is $p(i, j|\delta, \theta)$ can be defined as the joint probability of occurrence of gray levels i and j with a distance of δ and an angle of θ. Values of θ are usually $0°, 45°, 90°$ and $135°$. The image is scanned in these directions and the co-occurrences of gray-level is accumulated in the matrix. Matrix will have the size $N_G \times N_G$. The calculated histogram shows a more uniform spread for a fine texture than a coarse texture. To achieve the rotation invariance the PDF along all the directions are combined as in Equation 3.6. However, this increases the computational complexity of the problem.

$$p(i, j|\delta) = \sum_{\theta=0°,45°,90°,135°} p(i, j|\delta, \theta) \tag{3.6}$$

Therefore, GLCM has the information about the pixels in the image which are having similar gray levels. For an image with no texture, the matrix will only have diagonal values. When the image intensity variation increases other values of the matrix starts to increase too. A large set of features can be calculated on GLCM. According to Haralick et al. [91], 14 different features such as energy, variance, entropy, contrast, correlation, homogeneity, inverse difference moment, angular second moment, sum variance and sum entropy can be defined based on the GLCM. Among them, the most commonly used

features in previous literature, contrast, entropy, correlation and angular second moment
are computed as follows:

$$Contrast = \sum_{k=0}^{N_G-1} k^2 \sum_{i=1}^{N_G} \sum_{i=1}^{N_G} p(i,j|\delta); \qquad |i-j| = k \qquad (3.7)$$

$$Entropy = \sum_{i=1}^{N_G} \sum_{i=1}^{N_G} p(i,j|\delta) log p(i,j|\delta) \qquad (3.8)$$

if the mean and standard deviation of the image, in x and y directions are μ_x, σ_x and
μ_y, σ_y;

$$Correlation = \frac{\sum_{i=1}^{N_G} \sum_{i=1}^{N_G} i.j p(i,j|\delta) - \mu_x \mu_y}{\sigma_x \sigma_y} \qquad (3.9)$$

$$Angular\ Second\ Moment = \sum_{i=1}^{N_G} \sum_{i=1}^{N_G} p(i,j|\delta)^2 \qquad (3.10)$$

Gray-level difference matrix: In GLDM, first-order statistics based on the absolute gray-
level difference was computed between two pixels of the image separate by distance δ and
angle θ. Let's say $I(x,y)$ is the intensity image function and for a particular displacement
$\delta = (\Delta x, \Delta y)$ where Δx and Δy are integers, let $I_\delta(x,y) = |I(x,y) - I(x + \Delta x, y + \Delta y)|$. The
estimated PDF $p(i|\delta)$ of $I_\delta(x,y)$ is the vector having a number of dimensions equal to the
number of gray levels N_G in $I(x,y)$ and each component of the vector gives the probabil-
ity that i^{th} gray-level appears in $I_\delta(x,y)$.

When the image has a relatively coarse texture, then the displacement δ is small com-
pared to the size of the texture elements and therefore two points separated by a distance
δ will most probably have the same gray value. Therefore, $I_\delta(x,y)$ would be very small
and the PDF will be concentrated near zeroth gray-level. However, for a finer texture,
δ is larger than the texture size and $I_\delta(x,y)$ will have values in a larger range, giving a
broader PDF.

To analyse a directional texture, the amount of spread in the PDF should be calcu-
lated for various directions of δ. Similar to GLCM, usual angle values $0°, 45°, 90°$ and $135°$
are used to calculate the PDFs. The five most common properties used for analysis are
contrast, entropy, mean, angular second moment and inverse difference moment. These

properties are calculated based on the PDFs as follows:

$$Contrast\ (CONT) = \sum_{i=0}^{N_G-1} i^2 p(i \mid \delta) \tag{3.11}$$

$$Entropy\ (ENTR) = \sum_{i=0}^{N_G-1} p(i \mid \delta) \log p(i \mid \delta) \tag{3.12}$$

$$Mean\ (MV) = \sum_{i=0}^{N_G-1} i p(i \mid \delta) \tag{3.13}$$

$$Angular\ Second\ Moment\ (MV) = \sum_{i=0}^{N_G-1} p(i \mid \delta)^2 \tag{3.14}$$

$$Inverse\ Difference\ Moment\ (IDM) = \sum_{i=0}^{N_G-1} \frac{p(i \mid \delta)}{i^2 + 1} \tag{3.15}$$

Higher order statistics

Higher order statistics calculates the pixel relationships between three or more pixels. Gray-level run-length matrix (GLRLM) is an example of this type of method. These methods are less sensitive to noise.

Gray-level run-length matrix: Here the number of gray-level runs of different lengths are calculated from the image. Gray-level run means the number of consecutive pixels having the same gray value. The run length is number of pixels included within the gray-level run.

Therefore, for a fine texture, there will be more short gray-level runs and on the other hand a coarse texture will have longer gray-level runs. For an image $I(x,y)$ the GLRLM, $p(i,j|\theta)$ gives the number of times that gray-level i with run length j in the θ direction appears in the image. As with GLCM and GLDM, the common values for θ here are $0°, 45°, 90°$ and $135°$. Haralick et al. [82] showed the various run-length features that can be derived from the GLRLM, such as short run emphasis, long run emphasis, gray-level non-uniformity, run-length non-uniformity and run percentage.

If the maximum run length is R_{max}, the total number of runs is N_R and total number

of pixels in image is N_P, these features can be calculated as follows.

$$Short\ Run\ Emphasis = \frac{1}{N_R} \sum_{i=0}^{N_G} \sum_{j=0}^{R_{max}} \frac{p(i,j)}{j^2} \tag{3.16}$$

$$Long\ Run\ Emphasis = \frac{1}{N_R} \sum_{i=0}^{N_G} \sum_{j=0}^{R_{max}} p(i,j).j^2 \tag{3.17}$$

$$\text{Gray-level Non-uniformity} = \frac{1}{N_R} \sum_{i=0}^{N_G} (\sum_{j=0}^{R_{max}} p(i,j))^2 \tag{3.18}$$

$$\text{Run-length Non-uniformity} = \frac{1}{N_R} \sum_{i=0}^{R_{max}} (\sum_{j=0}^{N_G} p(i,j))^2 \tag{3.19}$$

$$Run\ Percentage = \frac{N_R}{N_P} \tag{3.20}$$

3.3 Related Work

Over the last three decades, various methods have been published to segment the airways from medical images. However, to the best of our knowledge, no known research has been carried out to segment the vocal fold area from CT images. Among the existing segmentation methods for airways, region growing methods [35], [37], [38], [39], machine learning methods [49], [50] and shape analysis methods [63], [66] are more popular. Selecting the most appropriate features is the most important part of any segmentation process. Many published papers have used features like appearance, shape and texture to detect the airway in different imaging modalities.

Computer vision-based texture analysis was first used in analysing information in geographical and satellite images [91], [93]. Also, textural features have been used as a successful discriminative feature for organs and tissue identification and pathological organ detection in medical images by many existing literature [94], [95]. In 1972, Suttan et al. [96] used textural information in medical images to automatically detect pulmonary disease in human lungs. Using three texture measures in lung X-ray images they reported 84% accuracy in differentiating abnormal lungs from normal lungs. Then, Chien

et al. [97] also demonstrated the importance of the texture features in classification of abnormalities in lungs using X-ray images of coal workers. Similar type of work was carried out by Hall et al. in 1974 [98] and Turner et al. in 1976 [99].

In 1982, [100] investigated the capability of detecting cirrhosis for liver CT images using power spectral density as a measure to analyse the texture in certain regions of the liver. However, they have not had any encouraging success in these experiments. Since 1985, many people have used textural analysis in ultrasound images. First, Raeth et al. [101] have used texture information in ultrasound images to differentiate pathological liver conditions like hepatitis, cirrhosis and fatty infiltration. Then in 1986, [102] used second order information derived from ultrasound scans to automatically cluster the liver tissues. Similar type of research was carried out by Garra et al. [103] in 1989 for classification of liver tissues in ultrasound scans using texture features. In 1992, [104] derived some texture features like gray-level difference statistics (GLDM), Fourier spectrum and spatial gray-level dependence matrices (SSGLDM) and used those features to classify ultrasound liver images using a Bayes classifier. They reported 90% successful classification for their data-set and suggested importance of texture features to analyse ultrasound images.

In 1993, Lerski et al. [105] used second order parameters like contrast, entropy, angular second moment and correlation derived from GLCM of brain MRI images to discriminate tumours, grey matter and oedema in human brain. In 1995, Mir et al. [94] used abdominal CT images of the liver region to show the capability of textural features to provide diagnostic information important to differentiate abnormal liver tissues. They employed GLCM, GLDM and GLRLM to extract number of texture parameters from CT images which were reduced to 256 gray levels. The experiments were able to detect the onset of liver disease which was difficult to differentiate for human observers and the three parameters, local homogeneity, entropy and gray-level distribution, were reported as the most effective texture features for various tissue type classifications. In the same year, Ito et al. [106] applied run-length analysis (GLRLM) to assess the risk of fracture in trabecula using CT images. Koss et al. [107] used GLCM based texture features to segment seven abdominal organs automatically from CT or MRI images. They input the

computed first and second order statistical features to a neural network classifier and reported 79%-100% accuracy in segmentation.

Chabat et al. [108] proposed a automatic method to differentiate various types of lung diseases using the energy, maximum, entropy, homogeneity and contrast feature descriptors calculated on GLCM of lung CT images. The feature vector was inputted to a Bayesian classifier and achieved a 73.6% sensitivity and 91.2% specificity. In 2004, Raicu et al. [109] investigated the most discriminative texture features to differentiate human organs. They analysed the Haralick and run-length encoding texture descriptors in CT images of five different organs, heart, liver, kidney, spleen and backbone, and suggested that sum of mean and high gray run emphasis features provide the most discriminative properties. Xu et al. [110] extended the run-length encoding matrix to 3D mode and showed that volumetric CT data of chest and abdomen can be differentiated using the texture feature descriptors calculated using the 3D run-length matrix. In run-length encoding method, a run is defined as the number of consecutive pixels having the same gray-level. Xu et al. showed that using large number of gray levels would result in most of the runs having only a single pixel, hence suggested that the number of gray levels should be reduced. They used linear mapping to reduce the original 16-bit images to 32 gray levels.

Podda et al. [111] applied run-length encoding technique (GLRLM) to segment the thrombotic tissue in aorta which was not possible to differentiate using standard gray level-based segmentation techniques due to having similar gray values as the surrounding tissues. They derived 11 different feature descriptors from the run-length matrix and showed the gray-level depth reduction has a great impact on the discriminative capability of the run-length matrix-based features.

[95] provided a 3D extension to Haralick features computed based on the GLCM calculated for a small cubic region of the volumetric abdominal CT data. They used a set of eight Haralick features and successfully identified all the organs except the kidneys. The specific spatial arrangement of different tissues in kidneys makes it difficult to detect using Haralick features. However, the experiments accidentally distinguished tumour

tissue in liver and exhibited the potential of this method to be used as an automated diagnostic tool.

In 2010, [112] proposed the SSGLDM, an extension to the GLCM, by adding the texture information of different spectral bands. They also defined new feature descriptors based on SSGLDM which were used in prostate cancer detection experiments and reported an improved accuracy compared to the traditional GLCM method.

Gomez et al. [113] used GLCM with multiple gray-level quantisation levels to classify breast cancer tissues in breast ultrasound images. First, they averaged the feature descriptors of the same distance over all the orientations to reduce the dimensionality of the feature space and found out that this gave a negative impact over the performance. The number of quantisation levels did not have any influence on the results and the contrast and correlation computed on the GLCM with 90° and over five pixels distance had the best differentiation power for the breast lesions.

In 2014, Torheim et al. [114] used second order feature descriptors derived from GLCM of MRI images to classify cervical cancer status. Experiments have reported 70% accuracy in predicting treatment outcome of the patients.

3.4 Data Selection and Generation

3.4.1 Data Selection

Axial images of the 44 subjects of Parkinson-MONASH data-set were carefully investigated and the subjects with images that are significantly blurred or low in contrast were excluded from this experiment. Furthermore, in some subjects the lungs were visible in images due to scanning position changes. Those subjects were also excluded from the study. As a result, we ended up with a set of 4D images of 20 subjects (five females and fifteen males), that had an acceptable quality for the experiment. Among those 20 subjects, there were six control subjects, six patients and eight uncategorised subjects.

3.4.2   Training and Testing Datasets Preparation

Each subject of the above selected data-set was provided with above 50 volumes of data. The first 49 volumes of all 20 subjects were chosen as the testing set. From the remaining volumes, pre-labelled training data set was prepared by manually extracting 96×96 pixels airway images and non-airway images. We used 3591 airway images and 2908 non airway images that were randomly selected to cover all 20 subjects as the training set to train the classifier. The testing set was not used in any way for the development of the system.

3.4.3   Ground Truth Generation

The airway area boundaries of each image slice of all the subjects were manually marked and extracted by the experts to validate the airway segmentation step. Furthermore, the ground truth rima glottidis area values were also provided for each subject. These were calculated manually by expert radiologists and used for the validation process.

3.5   Our Method

Glottis located in the airway and the human airway can be observed on an axial CT image as a dark circular area. Due to the size, shape and the intensity features this area can be spotted easily in a CT image slice. Therefore, we used this texture difference to segment the human airway from axial CT images of a laryngeal CT volume. Extracted texture features from the CT images were fed in to a SVM classifier to get the results.

Once the airway region was narrowed down, the exact airway pixels were segmented by adapting an automatic 2D region growing technique. The seed point for this region growing step was automatically calculated using the known information. This gave us the airway area in each slice of the laryngeal volume.

Among several slices in each frame, a scheme for selecting the glottis plane is developed and the area of rima glottidis from the selected frame is calculated to generate the

time series.

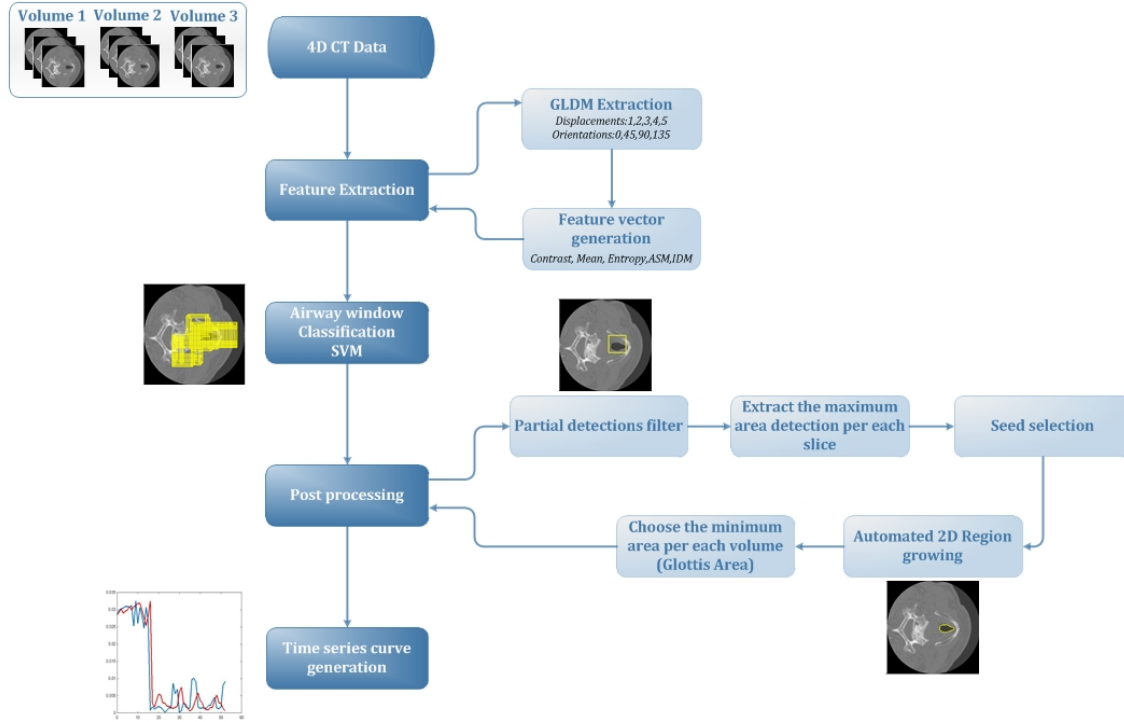The steps of the proposed methodology are shown in Figure 3.1.

Figure 3.1: Rima glottidis area detection framework

3.5.1 Feature Extraction

Due to the substantial gray-level difference in airway pixels and its neighbourhood, GLDM features [94] were used to extract texture for airway region. In GLDM, first order statistics based on the absolute gray-level difference was computed between two pixels of the image. To calculate the GLDM, the method explained in Section 3.2.1 were adapted, and to analyse the directional texture, the amount of spread in the PDF was calculated for various directions of δ. In this work, five different displacements ($\delta = 1, 2, 3, 4, 5$ *pixels*) in four possible forms $(0, \delta)$, $(-\delta, \delta)$, $(\delta, 0)$, $(-\delta, -\delta)$ which gave four directions $0°$, $45°$, $90°$ and $135°$, were used.

Furthermore, analysing the various feature descriptors proposed by Haralick et al. [91], five different properties were chosen as the feature descriptors. These feature descriptors

were calculated on the images according to the Equations 3.11-3.15. The description of the feature set used is shown in Table 3.1.

Table 3.1: Description of the feature set

Feature	Description
MV	Mean values of the GLDM probability density functions calculated for $1, 2, 3, 4, 5$ *pixels* distances in $0°$, $45°$, $90°$ and $135°$ directions (Equation 3.13)
CONT	Contrast of the GLDM probability density functions calculated for $1, 2, 3, 4, 5$ *pixels* distances in $0°$, $45°$, $90°$ and $135°$ directions (Equation 3.11)
ENTR	Entropy of the GLDM probability density functions calculated for $1, 2, 3, 4, 5$ *pixels* distances in $0°$, $45°$, $90°$ and $135°$ directions (Equation 3.12)
ASM	Angular Second Moment of the GLDM probability density functions calculated for $1, 2, 3, 4, 5$ *pixels* distances in $0°$, $45°$, $90°$ and $135°$ directions (Equation 3.14)
IDM	Inverse Difference Moment of the GLDM probability density functions calculated for $1, 2, 3, 4, 5$ *pixels* distances in $0°$, $45°$, $90°$ and $135°$ directions (Equation 3.15)

Airway region has the minimum intensity level with respect to the other regions in a CT image. Therefore, the mean intensity in that area should be the lowest. Also, inside the airway region two neighbouring pixels should have roughly the same intensity level. But in the boundary region the intensity shows a large variation from dark to light, hence the contrast is maximum.

Entropy will measure the randomness of the image. CT images normally have a random texture due to various absorption rates in soft tissues, bones and air. Furthermore, noise, artefacts, beam hardening and partial volume effect may also generate some randomness in the images. However, within the region of human airway we can expect some uniformity compared to the other regions, as it is an air-filled area.

Angular second moment gives an indication of the homogeneity or uniformity in the image. A homogeneous image normally has few gray levels, which means gray-level

difference will be lower most of the time. This gives few gray levels in the GLDM but relatively high values in the PDF, $p(i|\delta)$. Therefore, relatively uniform airway region windows give high angular second moments. Similarly, inverse difference moment is also a measure of local homogeneity of the image. According to the calculation in Equation 3.15 this will get a small contribution from the inhomogeneity regions thus the resulting inverse difference moment for inhomogeneity areas are low and homogeneity area are high. As the airway region can be considered homogeneous compared to the rest of the regions, that will give a fairly high inverse difference moment. So local entropy of that area should be comparatively low.

Therefore, mean, contrast, entropy, angular second moment and inverse difference moment were chosen as the feature descriptors for our data set. In this experiment, all possible displacements and angles gave 20 different PDFs and calculating these descriptors for each PDF resulted in 100 texture feature descriptors per image window. The feature extraction algorithm was implemented in Matlab 2014a Computer Vision toolbox.

Furthermore, feature vectors were normalised using z-score normalisation to bring them to the same scale, so that the features vectors on different scales will not influence the classification differently.

Image preparation

The images in the data-set were each 512×512 pixels and 12-bit 2D CT images. GLDMs are sensitive to the size of the texture samples. Therefore, using the original images as it is to calculate the GLDM will give matrices which are larger than the image. These large matrices always have lot of information about the very fine texture details. When such fine details are not a target of interest the images are often quantised to have fewer number of gray levels before calculating GLDMs. In 1995, Mira et al. [94] showed that reducing number of gray levels is sufficient to segment the texture details in medical images.

Human airway appears as a relatively larger, darker region on a 2D axial CT image slice (Figure 3.2) and hence the original resolution (4096 gray levels) will give lot of unwanted texture details which we do not require for the segmentation. Therefore, to save time and computational cost we quantized the images to 256 gray levels before calculating the GLDMs.
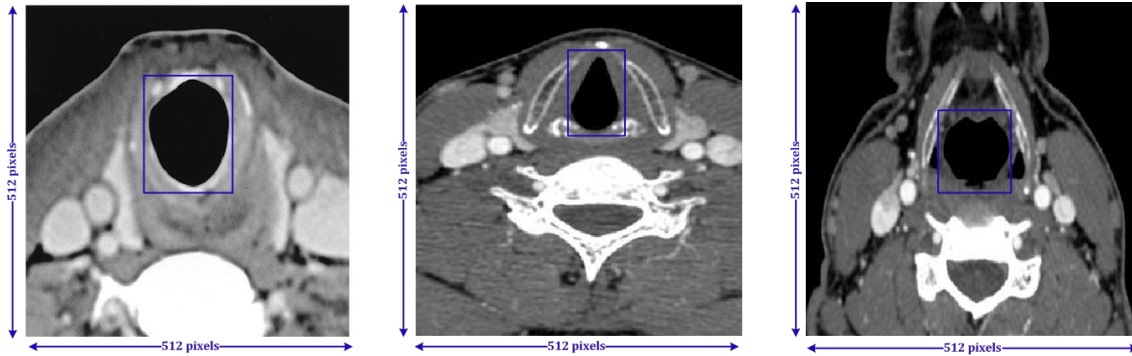

Figure 3.2: Airway size with respect to the axial image

Image acquisition

The sliding window technique was used to extract image samples from the original 512×512 pixels axial images. The size of an image sample was fixed to 96×96 pixels. This sample size was decided empirically by carefully observing the data-set to include the whole airway region inside.

Therefore, for airway detection 12 cells blocks were used, where the cell size was fixed at 8×8 pixels. The 512×512 pixels image was considered as a dense grid of uniformly spaced cells and the feature vectors were computed on each block (detection window). Blocks were extracted by sliding the detection window (Figure 3.3) with an 8-pixel (one cell) stride. This tight overlapping for extracting the detection windows was used for getting the minimum miss rate (MR) of an airway.
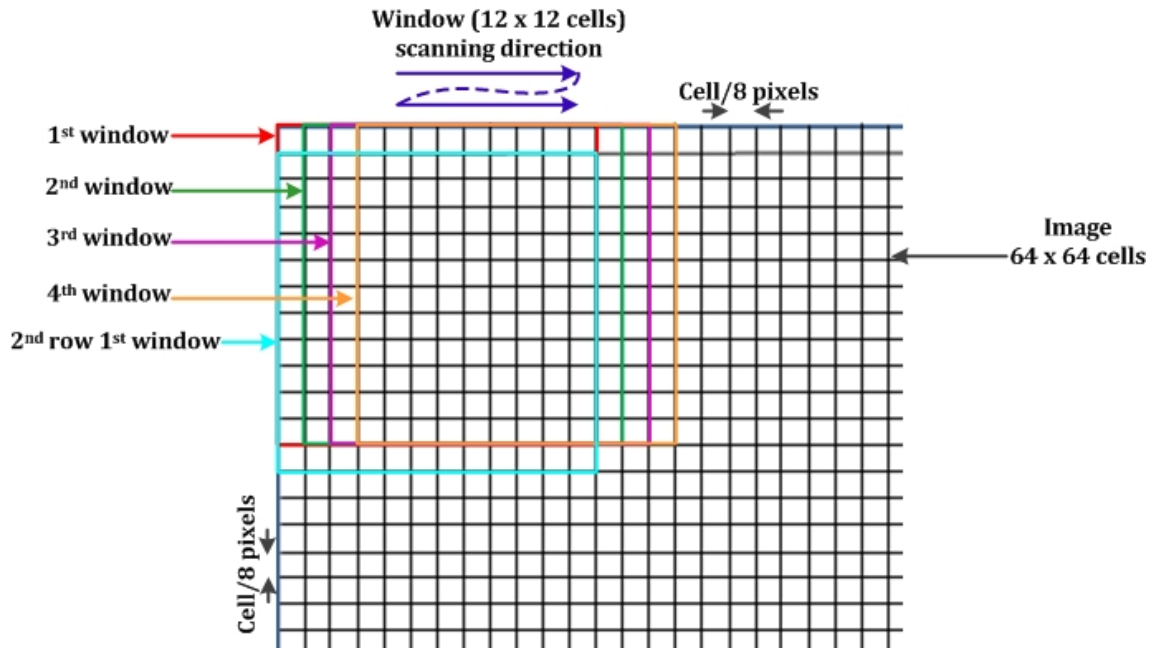
Figure 3.3: Sliding window technique

These acquired blocks/image windows was used to calculate the GLDM and then the 100 size feature vectors. Feature vectors of each block were normalised using their z-scores.

3.5.2 Classification

In Computer Vision, many different feature classification methods can be seen [115]. Due to the ability of performing better in higher dimensional spaces, which is often the case of texture analysis, SVM were chosen for this automatic classification.

Support vector machines

SVMs were introduced by Vapnik in 1999 [116] and are based on the principle of searching a hyper-plane that maximises the separation between two classes of features. SVMs are supervised classifiers which build a model by analysing the training data samples, which are labelled into one of the two categories. When an unlabelled new sample is given, the model classifies the sample into one category or other. This model maps the training samples into a space which they can be separated using a hyper-plane, which has the maximum distance from the support vectors (Figure 3.4). When a new data sample is

given, it is also mapped in to the same space and the class of that is predicted according to the side of the hyper-plane it falls on. Support vectors are the samples of the training data-set, which are on the boundary of both classes.
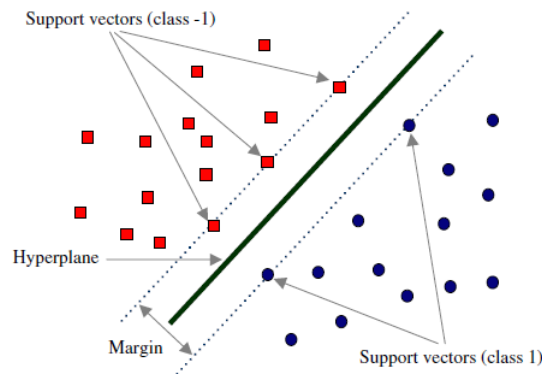

Figure 3.4: Maximum margin hyper-plane separation in support vector machines

In SVM classification it is an important common practice to separate the training set from the testing set, as if the same data-set is used to train the classifier then the resulting model will just repeat the class labels that it had just learn, but will not be able classify any unseen data. This problem is called over-fitting in classification and to avoid this the model should be learned on a separate training set each time.

When it is not possible to find a linear separation between the data points, a kernel trick is used to find a non-linear separation [117]. In this case, data points are non-linearly mapped to a higher dimensional feature space, so that the points are mostly linearly sep-arable in that feature space. The hyper-plane found in that feature space can be converted to a non-linear boundary in the original input space.

When training samples $X = x_1, x_2, x_3, ..., x_m$ are, a vector in the input space $I = R^n$ that each point belongs to one of the two classes. Vector $Y = y_1, y_2, y_3, ..., y_m$ is the vector de-notes the class index of the each point which can be $+1$ or -1. Training a non-linear SVM will map the input space data X into a higher-dimensional feature space $F = R^N$ using a non-linear mapping function ψ, where $z = \psi(x)$ and find the linear separation function f with weight w and bias b, in feature space F (Equation 3.21). Sign of $f(x)$ gives the class index of the points in input space. $f(x) = 0$ denotes the non-linear boundary in the input space.

$$f(x) = w.\psi(x) + a \tag{3.21}$$

If the data can be separated linearly, there are two parallel hyper-planes (H_1 and H_2) that divide the two classes. This H_1 and H_2 hyper-planes can be described by Equations 3.22 and 3.23. Support vectors are the data points that lie on these hyper-planes.

$$w.\psi(x) + a = 1 \tag{3.22}$$

$$w.\psi(x) + a = -1 \tag{3.23}$$

Margin is the distance between these two hyper-planes and geometrically this distance can be proven as $2/\|w\|$. The hyper-plane gives the maximum separation between the two classes and is located in the middle of H_1 and H_2 hyper-planes which is described by Equation 3.24.

$$w.\psi(x) + a = 0 \tag{3.24}$$

The distance between H_1 and H_2 ($2/\|w\|$) should be maximised to find the solution to the SVM. Hence $\|w\|$ should be minimised.

For correct classification of each input sample, all the points should satisfy Equations 3.25 and 3.26.

$$y_i = 1 => w.\psi(x_i) + a \geq 1 \tag{3.25}$$

$$y_i = -1 => w.\psi(x_i) + a \leq -1 \tag{3.26}$$

These two equations can be combined as follows.

$$y_i(w.\psi(x_i) + a) \geq 1 \quad for\ all\ 1 \leq i \leq m \tag{3.27}$$

To find the minimum $\|w\|$, the separating boundary equation should be solved (3.28).

$$y_i(w.\psi(x_i)) + a - 1 = 0 \tag{3.28}$$

According to Cristianini et al. [118], this optimisation problem to find the minimum $\|w\|$ can be expressed using the support vectors x_i as follows:

$$w = \sum_{i=1}^{m} \alpha_i y_i \psi(x_i) \quad where\ \alpha_i\ are\ positive\ real\ numbers\ and\ x_i\ are\ support\ vectors \tag{3.29}$$

Therefore, the classification decision function can be rewritten as,

$$f(x) = \sum_{i=1}^{m} \alpha_i y_i \langle \psi(x_i), \psi(x) \rangle + b \tag{3.30}$$

When working with higher dimensional feature spaces, computational complexity of Equation 3.30 becomes very high. To avoid this problem, we use the kernel trick in non-linear classification problems.

Kernel is defined as a function that gives the dot product of two vectors.

$$K(x_i, x_j) = \langle \psi(x_i), \psi(x_j) \rangle \tag{3.31}$$

Therefore, we can simply rewrite the decision function using the kernels as follows.

$$f(x) = \sum_{i=1}^{m} \alpha_i y_i K(x_i, x) + b \tag{3.32}$$

Using kernels in the decision function to avoid the requirement of explicitly calculating the dot products between the data samples in the feature space, hence makes it possible to map to a very high dimensional feature spaces. Commonly used kernels are polynomial kernels, Gaussian (radial basis function (RBF)) kernels and sigmoid kernels. Each of these kernels need some parameters to be set by the user. Normally, this is done by doing a grid search on all possible kernel and parameter values and evaluating the performances. Some prior knowledge about the problem will always be useful. Therefore, the classification performance of SVM always depends on the separability of the data, kernel and the parameters chosen.

When choosing the optimal kernel and the parameters manually, there is a risk of over-fitting the model as one can adjust the kernel and parameters until the model gives the optimal performance. This will allow some knowledge about the testing set to leak to the model and may cause over-fitting. To avoid this, there should be a separate set for validation as well. However, keeping three different data-sets (training, validation and testing) will substantially reduce the data available for training. To overcome this issue, a technique called cross-validation is performed on the training set. In cross-validation a separate validation set is not required. Instead, if $n - fold$ cross-validation is performed, the training set is divided into n number of smaller sets. Then for each n folds Algorithm 1 is performed.

Algorithm 1 : SVM $n - fold$ cross-validation algorithm

    STEP 1 : The SVM model is trained using $(n - 1)$ smaller sets as the training set

    STEP 2 : The remaining set is used to validate the resulting model

    STEP 3 : The average accuracy of all the folds is reported as the accuracy of that particular kernel and particular parameters

---

Although the computational complexity of Algorithm 1 is high, this will avoid the wastage of data.

Although SVM classifiers are normally used for binary classification, these can be extended to solve multi-class problems as well. In that case, one-against-all technique is adapted. This uses a series of binary SVM classifiers and each classifier is trained to separate one class from all the others. As an example, if there are N classes, N trained SVM classifiers are used and the kth classifier is trained to separate class k from other (N-1) classes. These multi-class SVMs are popular as they are easy to implement and fast to train and predict.

One drawback of SVM classification being a supervised learning model is that it requires pre-labelled training data samples which may be difficult to find in a real world problem. For this unsupervised learning model, support vector clustering is used. SVM is improved using a clustering algorithm, and in large-scale industrial applications SVM

clustering is used as a pre-processing step before the classification, when no labelled data has been provided.

Selection of Kernel and parameters

In this work, SVM was implemented in Matlab 2014a using LibSVM library [119].

Feature vectors calculated on 3591 airway images and 2908 non-airway images in the training data-set were used to train the SVM. We performed 10-fold cross-validation and a grid search to decide the optimal kernel and kernel parameters; C the parameter for soft margin cost function and γ the free parameter of Gaussian kernel. Linear, polynomial and RBF kernels were chosen as options and the precision, recall, f-Score and accuracy were evaluated for parameter values between 2^{-20} and 2^{20}. The maximum performance given by the three kernels are listed in Table 3.2. Therefore, RBF kernel with $c = 2^{5.5}$ and $\gamma = 2^{-6.5}$ parameters was chosen as the kernel that gives the optimum performance.

Table 3.2: Results of classification using a 10-fold cross-validation

Kernel	C	γ	Precision (%)	Recall %	f- Score	Accuracy (%)
Linear	2^{10}	2	99.28	82.30	0.90	99.87
Polynomial	2^{20}	2^{-8}	99.32	87.03	0.93	99.82
RBF	$2^{5.5}$	$2^{-6.5}$	99.44	87.45	0.93	99.86

Prediction of results

The SVM was trained using the above selected optimal kernel and parameters and the features calculated on the total training data-set (3591 airway images and 2908 non-airway images). Then, the trained model was used to classify the texture features calculated on the testing set of 4D image data of 20 subjects.

3.5.3 Removal of Redundancies and False Positives

Even after thresholding by the predicted probabilities, the output of the SVM still had some FPs mostly in the form of partial detections. Moreover, as we used overlapping sliding window technique to extract the input image patches to the SVM, predictions also contained a lot of overlapping detections which are redundant. To discard these false detections and redundancies a threshold-based technique was applied. Two threshold values were used in this experiment, T_{min} and T_S, which define the gray levels of background areas and air-filled areas, respectively.

During processing, the background pixels were not taken into consideration. Therefore, as the first step background pixels were removed from all the detected windows using the threshold T_{min}. Then the partial detections were removed to filter only the windows with whole airway inside. For this, the four boundaries of each window were analysed and any window having boundary pixels with gray-level smaller than T_S was discarded. Then for each remaining window, the number of pixels having gray levels smaller than the threshold T_S were calculated and the maximum window was taken as the optimal airway window for the respective image slice.

The two threshold values, T_{min} and T_S were calculated automatically for the data-set using the below mentioned methods.

Calculation of threshold T_{min}

To calculate the threshold gray value to remove the background region from the CT image, Algorithm 2 was used. For the data-set used in this experiment, the value of the T_{min} was -2048.

Calculation of threshold T_S

To calculate the threshold value T_S, optimal threshold method (Otsu's method) [120], [121] was used. Optimal threshold method automatically separates two classes of pixels in an image and calculates the optimal threshold, so that the intra-class variance is minimal.

Algorithm 2 : Algorithm to obtain the background threshold intensity of a 2D image slice

Input: $I(x, y)$ ▷ 512×512 2D axial CT image slice
Output: T_{min} ▷ background threshold intensity
 1: $T_{min} = 0$ ▷ initialise the threshold to zero
 2: **for** $x = 1$ to 512 **do**
 3:   **for** $y = 1$ to 512 **do**
 4:     **if** $(x = 1$ **or** $x = 512$ **or** $y = 1$ **or** $y = 512)$ **and** $(I(x, y) < T_{min})$ **then**
 5:       $T_{min} = I(x, y)$ ▷ update the threshold to new minimum value
 6:     **end if**
 7:   **end for**
 8: **end for**
 9: **return** T_{min}

For this, first the background pixels were removed from images using the threshold T_{min} calculated in the above step. Then we assumed that the resulting image volume of larynx contains only two types of voxels.

1. *Body voxels*: Voxels within the body of the subject having higher Hounsfield values

2. *Non-body voxels*: Voxels within the airway or other air-filled parts of the body having lower Hounsfield values

Based on this assumption we executed the Algorithm 3 to calculate the T_S.

Algorithm 3 requires an initial threshold as an input. According to Wu et al, [122], air in a CT image will appear with the mean Hounsfield value of $-1000HU$ and airway, tracheal and lung voxels will have a value between $-910HU$ to $-500HU$. Bones, cartilages, blood and other body voxels will have values well above $-500HU$. Therefore, the initial threshold T_0 was calculated based on these values. However, in our data-set, for any initial threshold between $0HU$ to $-1000HU$, the calculated T_S from the above optimal thresholding technique was $-436HU$.

3.5.4   Region Growing

Region growing is a pixel or region-based image segmentation method which uses an initial seed point as an input. The method then analyses the connected neighbourhood

Algorithm 3 : Algorithm to obtain the optimal threshold intensity for air-filled areas in a 2D CT image slice

---

Input: $I(x, y)$ ▷ 512×512 2D axial CT image slice
      T_{min} ▷ threshold intensity of background
Output: T_S ▷ optimal threshold intensity for air-filled areas
  1: $T_S = -1000$ ▷ initialise the threshold to the mean Hounsfield value of air in a CT image
  2: **for all** x, y such that $1 \leq x \leq 512$ **and** $1 \leq y \leq 512$ **do**
  3:    $(I(x, y) \leq T_{min}) = 0$ ▷ remove the background
  4: **end for**
  5: **repeat**
  6:    $T_0 = T_S$ ▷ update the current threshold T_0
  7:    **for all** x, y such that $1 \leq x \leq 512$ **and** $1 \leq y \leq 512$ **do**
  8:    $\mu_{NB} = \dfrac{\sum_{1}^{N_{NB}} (I(x,y) \leq T_0)}{N_{NB}}$ ▷ mean of the *Non-body voxels* (N_{NB}: Number of *Non-body voxels*)
  9:    $\mu_{B} = \dfrac{\sum_{1}^{N_{B}} (I(x,y) > T_0)}{N_{B}}$ ▷ mean of the *Body voxels* (N_{NB}: Number of *Body voxels*)
10:    $T_S = \frac{\mu_B + \mu_{NB}}{2}$ ▷ calculate the new threshold
11:    **end for**
12: **until** $(T_S - T_0) \leq 1$
13: **return** T_S

---

pixels and decides whether they should be added to the region or not depending on the membership criterion. Pixel intensity, texture or colour could be used as the membership criterion. As an example, if pixel intensity is chosen as the membership criterion, which is the usual case in most cases, a pixel intensity threshold is used to determine the insertion criteria to the region. This is an iterative process which continues until there are no connected pixels in the neighbourhood or none of the pixels matches with the membership criterion. Choosing the suitable threshold value is important to get the desired results. This can be done using the knowledge about the problem or examining the histogram of the image.

Neighbourhood can be scanned in two different ways from the seed pixel, as 4- or 8-connected neighbourhood (Figure 3.5). Scanning starts from the seed point. This point can be chosen based on the user requirements, such as a pixel with a certain gray-level or a pixel in a particular location. However, the seed point is important as the final results are dependent on this initial input. This is one problem in region growing, as this input is mostly manual and depends on the user. In our experiment, we have automated the seed selection algorithm as well.

Region growing is a fairly simple concept and needs only few inputs, the seed point and the threshold. However, as this is a pixel-based iterative process this could be computationally expensive. When the image has sharply defined edges the region growing provides better segmentation results. However, this method is sensitive to noise.
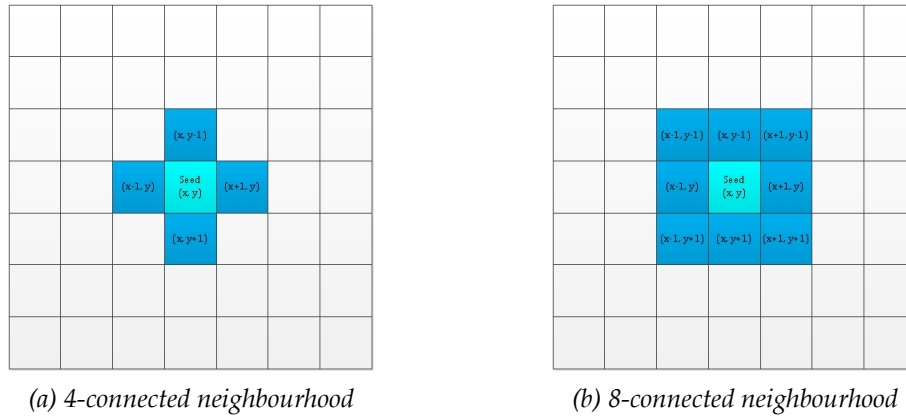

(a) 4-connected neighbourhood   (b) 8-connected neighbourhood

Figure 3.5: Different scanning ways of pixel neighbourhood

In this experiment, the resulting optimal airway windows after the FPs and redundancies removal step (Section 3.5.3), contains a clearly defined airway boundary due to the substantial intensity difference in the airway and surrounding soft tissues. Therefore, to find the airway pixels a fully automated 2D region growing technique was used. The region growing algorithm implemented was comparable to the algorithm proposed by Van et al. in [27].

The algorithm was applied to each axial slice separately. The seed point for each image slice was calculated automatically using the gray-scale information we have. First, all the background pixels (pixels having gray value T_{min} or less) were excluded from the image slice . Then the pixel having the lowest gray value was taken as the seed for that slice. From that seed, the area of the airway was grown using the same threshold value T_S calculated in the above 3.5.3 step. The complete steps are illustrated in Algorithm 4.

Algorithm 4 : Region growing algorithm to obtain the pixels of the airway region in a 2D CT image slice

---

Input: $W(x,y)$ ▷ 96×96 airway window from the 2D axial CT image slice
      T_{min} ▷ threshold intensity of background
      T_S ▷ optimal threshold intensity of airway region
Output: P ▷ a set of points of the airway region
      N_{seg} ▷ number of segmented airway pixels

  1: **for all** x, y such that $1 \leq x \leq 96$ **and** $1 \leq y \leq 96$ **do**
  2:   $(W(x,y) \leq T_{min}) = 0$ ▷ remove the background
  3: **end for**
  4: $C_0 \leftarrow find(\min(W(x,y)))$ ▷ seed point $C_0 = (x_0, y_0)$
  5: $I_0 = W(x_0, y_0)$ ▷ seed point intensity value
  6: $C_{curr} = C_0$ ▷ current minimum intensity pixel is seed point
  7: $Push(C_0, P)$ ▷ initialise airway pixels list with seed point
  8: $Push(C_0, L_{neigh})$ ▷ initialise neighbourhood list with seed point
  9: $I_{mean} = I_0$ ▷ initialise segmented pixel mean to seed point intensity
10: $I_{curr} = I_0$ ▷ initialise current pixel intensity to seed point intensity
11: $N_{seg} = 1$ ▷ initialise number of segmented pixels (only seed point)
12: $N_{all} = 96 \times 96$ ▷ number of pixels in image window
13: **while** $((I_{curr} \leq T_S)$ **and** $(N_{seg} < N_{all}))$ **do**
14:   **for all** $C_i = (x_i, y_i)$ in *4-connected neighbourhood* **of** C_{curr} **do**
15:     **if** $(W(x_i, y_i) \leq T_S)$ **and** $C_i \in W(x,y)$ **and** $(C_i \notin P)$ **and** $(C_i \notin L_{neigh})$ **then**
16:       $Push(C_i, L_{neigh})$
17:     **end if**
18:   **end for**
19:   $C_{curr} \leftarrow find(\min(|L_{neigh} - I_{mean}|))$ ▷ segmented pixel $C_{curr} = (x_{curr}, y_{curr})$
20:   $I_{curr} = W(x_{curr}, y_{curr})$
21:   $Push(C_{curr}, P)$
22:   $N_{seg} \leftarrow N_{seg} + 1$
23:   $I_{mean} = mean(P)$
24: **end while**
25: **return** P, N_{seg}

3.5.5 Calculation of Area and Generation of Temporal Maps

Once the region was accurately segmented using the region growing method, the areas were generated by using the voxel information in the original NifTI-1 header and the number of pixels in the region. The equations used to calculate the area and volume are shown in Equations 3.33 and 3.34.

$$Area = \frac{Number\ of}{pixels\ in\ Slice} \times \frac{Voxel}{Width} \times \frac{Voxel}{Height} \tag{3.33}$$

$$Volume = \frac{Total\ number\ of}{pixels\ in\ Subject} \times \frac{Voxel}{Width} \times \frac{Voxel}{Height} \times \frac{Slice}{Thickness} \tag{3.34}$$

From the detected airway images, the rima glottidis plane was identified using the known anatomical features of the human larynx. According to the human laryngeal anatomy, the rima glottidis is the narrowest part of the laryngeal cavity. Therefore, it can be assumed that the slice with the minimum airway area is the rima glottidis plane.

However, to allow a significant tolerance to minimise errors, an N number of slices around the slice with the minimum area were considered. The average area of all those slices ($2N + 1$ slices) were taken to generate the temporal maps.

All the values were normalised with respect to the subject before plotting the temporal maps. Then the similarity indexes, correlation coefficient, mean squared error, Euclidean distance and dynamic time warping distance were measured between the generated graph and the ground truth graph for each subject.

Choosing a value for N was done using a variable slice analysis technique. Time series similarity indices were calculated for each subject, using different values of N ranging from 0 to 10. Evaluating this result, it was understood that subjects performed well when N was 2. Therefore, to generate the time series plots an average area of five slices around the minimum area slice was taken in this experiment.

3.6   Results and Performance Validation

3.6.1   Results

Performing the region growing on post-processed classifier output gives the airway delineation for all the patients. Samples of those results, which were taken from frame 1 of each subject, have been shown in Figure 3.6.
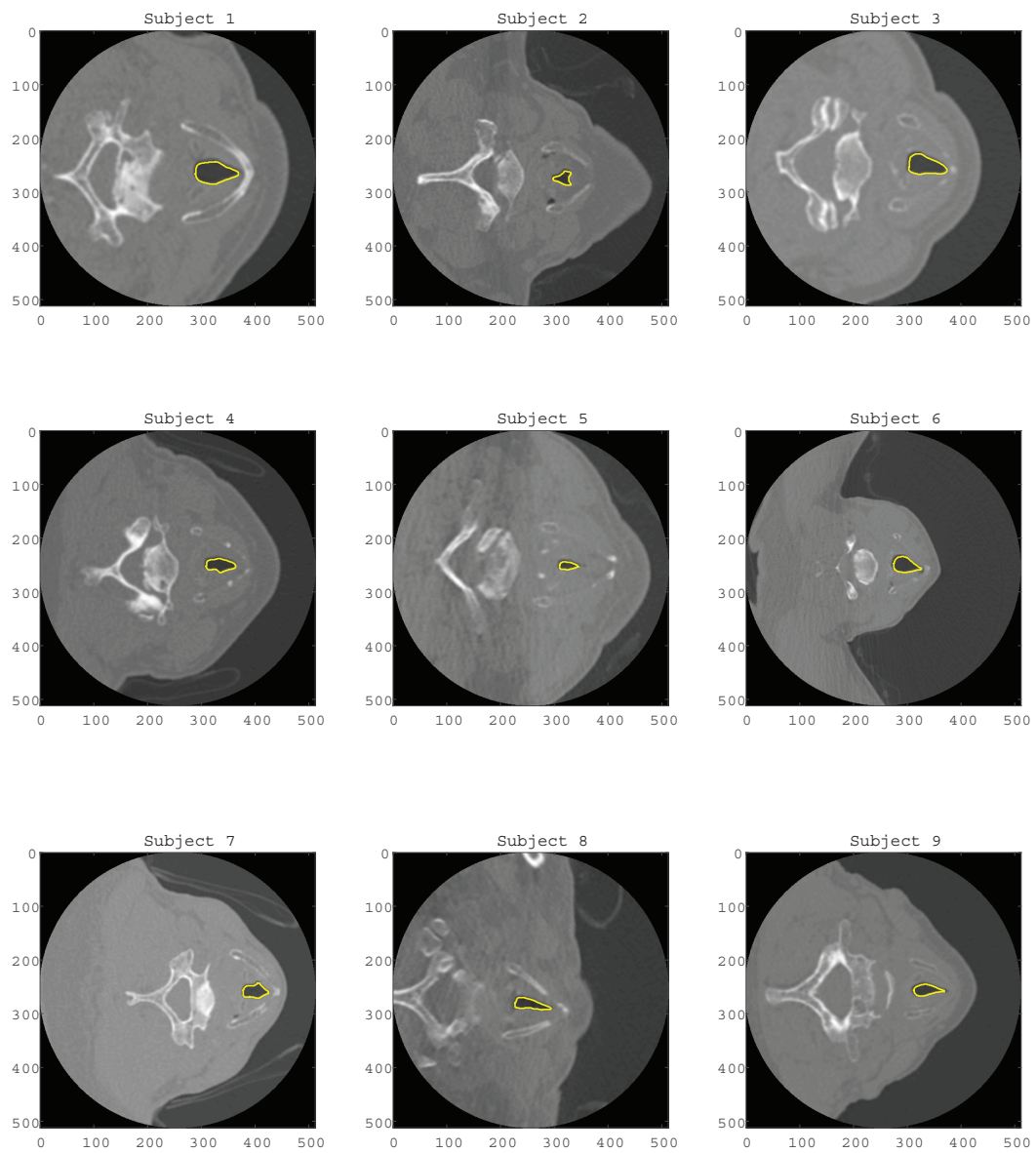
Figure 3.6: Airway delineation output from region growing step
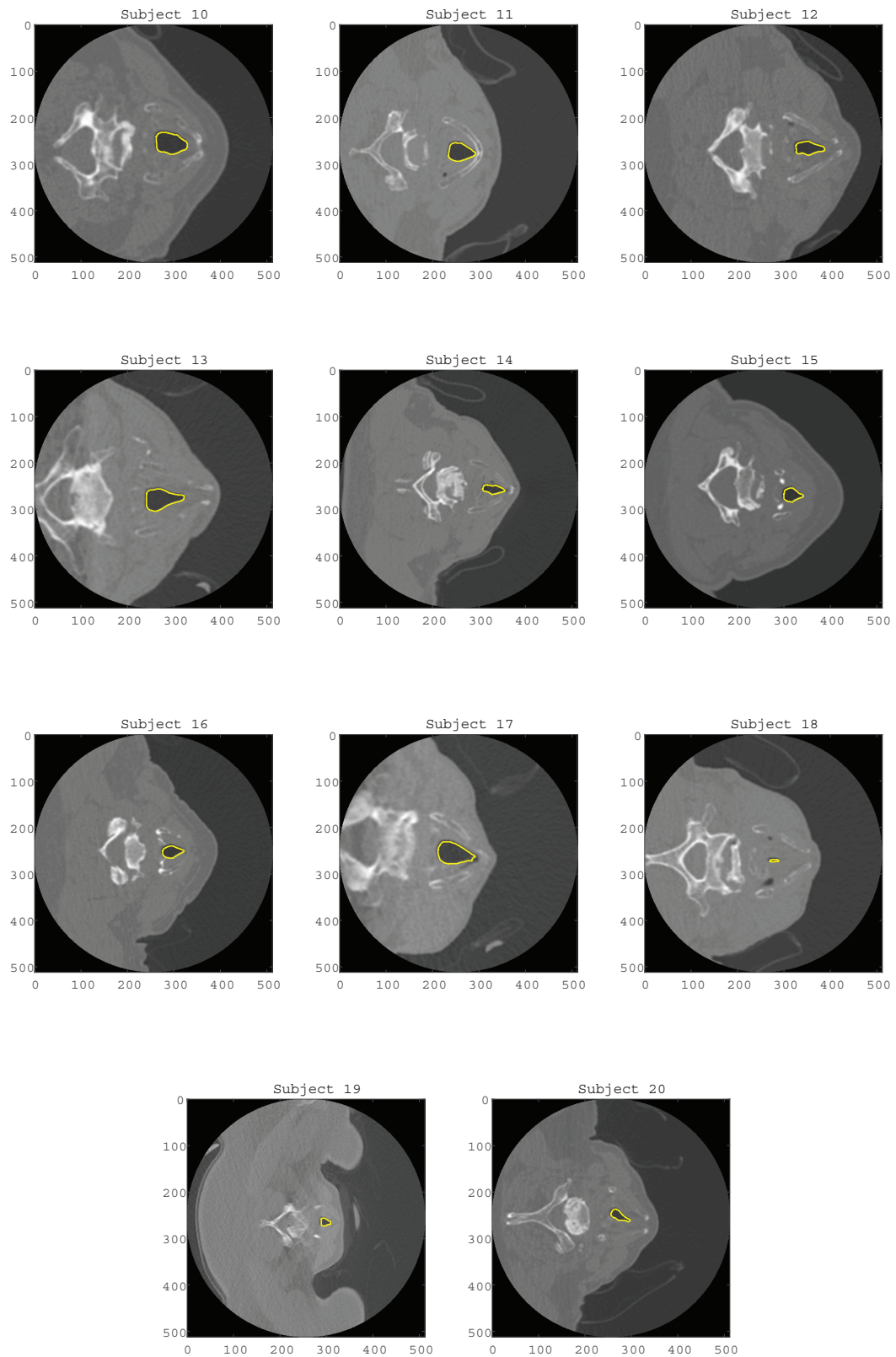
Figure 3.6: Airway delineation output from region growing step

Furthermore, the temporal maps drawn using the detected rima glottidis area of each frame for all controls, patients and uncategorised subjects are shown in Figures 3.7, 3.8 and 3.9, respectively.
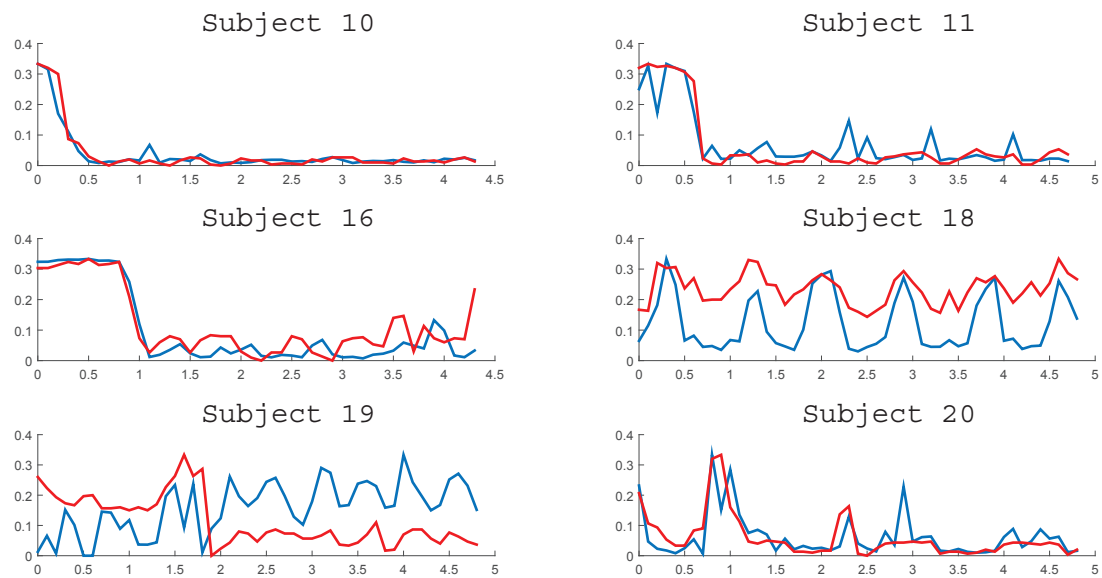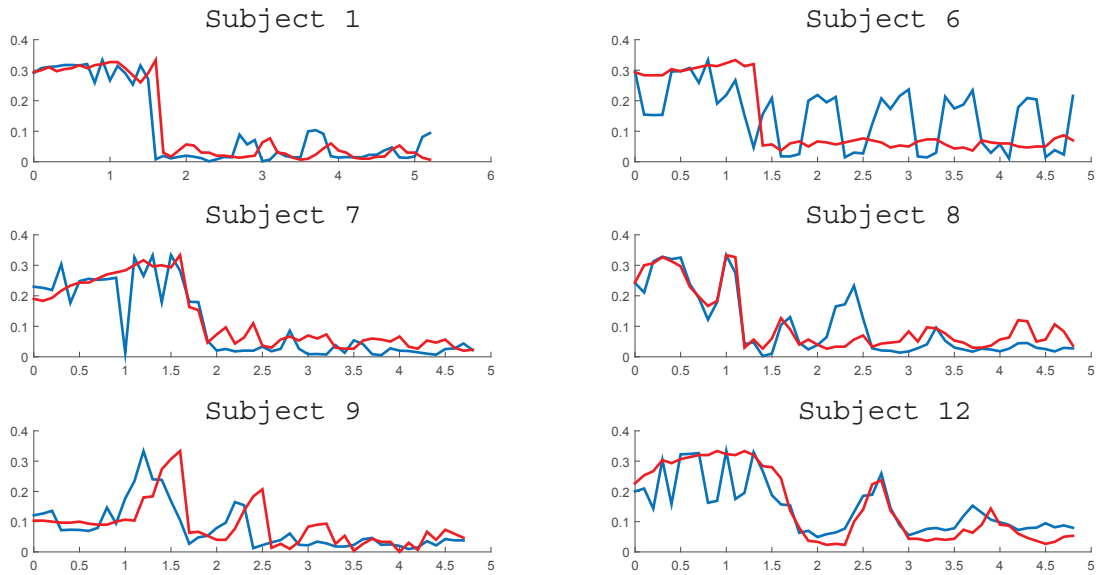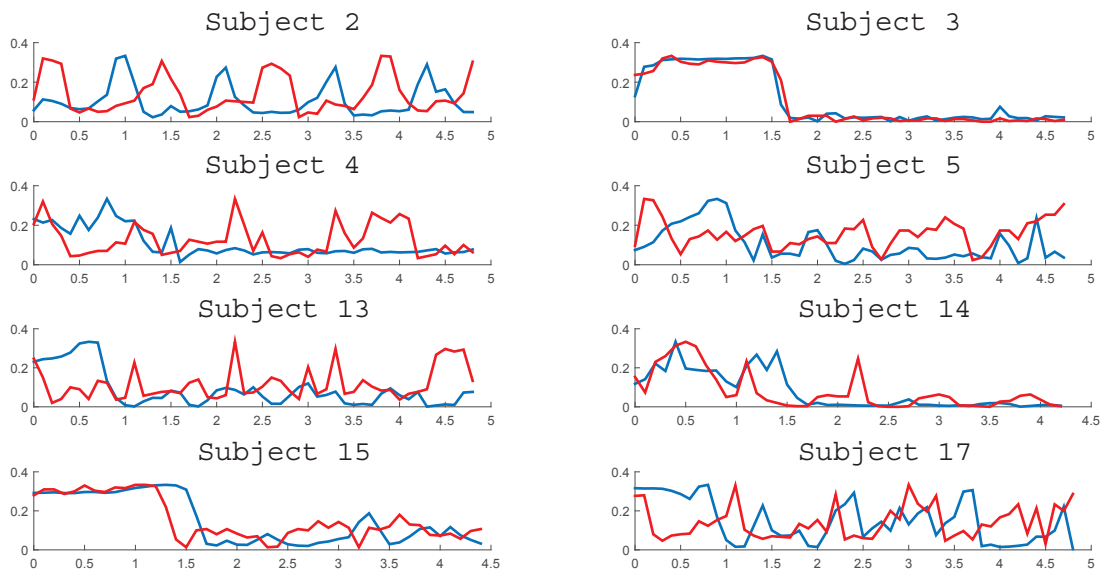

Figure 3.7: Time series of rima glottidis area for control subjects. Red indicates ground truth and blue indicates calculated values. Horizontal axis (time/s), Vertical axis (Normalized area)

Figure 3.8: Time series of rima glottidis area for PD patients.Red indicates ground truth and blue indicates calculated values. Horizontal axis (time/s), Vertical axis (Normalized area)

Figure 3.9: Time series of rima glottidis area for uncategorised subjects.Red indicates ground truth and blue indicates calculated values. Horizontal axis (time/s), Vertical axis (Normalized area)

3.6.2 Performance Validation

SVM classifier performance

The performance of the airway window classification of the svm classier was analysed using the ground truth data of the airway region. Ground truth annotations for the airway region were provided by the expert radiologists in the Monash Movement Disorder Clinic. The class labels for the test image windows were generated using these annotations. A 96×96 window was considered as a true positive (TP) when the whole airway region is inside the window. Otherwise, the windows were taken as a FP. The prediction probabilities of the test image windows of the SVM were compared with these class labels using different thresholds to generate the receiver operator characteristic (ROC) curve.

ROC curve drawn between $sensitivity$ and $1 - specificity$ for the different thresholds are shown in Figure 3.10. Area under the ROC curve was reported as 0.8503. The best $specificity$ and $sensitivity$ values calculated based on the ROC curve were 0.8754 and 8252 respectively (shown with a red marker on the ROC curve in Figure 3.10). The FPR corresponding to this best performed threshold was 0.1246. This means, despite the significant number of FPs, our SVM model classifies the airway windows and non-airway windows accurately.
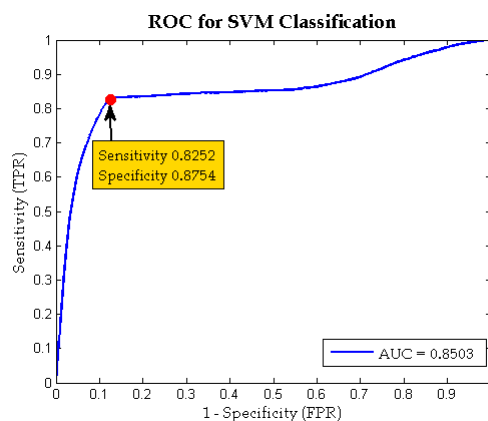

Figure 3.10: Receiver Operating Characteristic curve for the support vector machine classification

SVM classifier performance after false positives and redundancies removal

To improve the classification accuracy of the SVM, we employed a FPs and redundancies removal step on the SVM results. First, the SVM prediction probabilities were filtered using a hard threshold of 0.9 to obtain a higher accuracy. Then the resulting windows were analysed and partial detection and overlapping detections were removed. MR, precision (PRE), sensitivity/recall (SEN), specificity (SPE), FPR, false positives per image slice(FPPI) and f-score were calculated on the outcome. TP, FP, true negatives (TN) and false negatives (FN) computed on the outcome were used in the below equations (3.35-3.40).

$$MR = \frac{FN}{TP + FN} \tag{3.35}$$

$$FPR = \frac{FP}{FP + TN} \tag{3.36}$$

$$PRE = \frac{TP}{TP + FP} \tag{3.37}$$

$$SEN = \frac{TP}{TP + FN} \tag{3.38}$$

$$SPE = \frac{TN}{TN + FP} \tag{3.39}$$

$$FPPI = \frac{FP}{Total\ Image\ Slices} \tag{3.40}$$

$$f - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3.41}$$

The results are shown in Table 3.3.

Table 3.3: Results of the classification performance after post processing for all the subjects

MR	FPPI	FPR	PRE	SEN	SPE	f-score
0.1304	0.0633	0.0001	0.9279	0.8696	0.9999	0.8978

According to the Table 3.3, our SVM classifier has a 0.1304 MR with 0.0633 FPPI slice. FPR was 0.0001, which was very low compared to the FPR of SVM model alone and hence shows the effectiveness of this FPs and redundancies removal step.

However, number of negatives in our test data were very high compared to the number of actual airway windows. This is the natural phenomenon in laryngeal CT data as each axial image slice only consists of one airway window. The extremely low FPR and the very high specificity are driven by this reason and hence less informative in deciding the model prediction accuracy. However, the high precision, sensitivity and f-score indicates that the accuracy of our model is substantial.

Region Growing Segmentation Performance

The output of the region growing segmentation was compared with the ground truth segmentations of the airway window areas. TP, FP, TN and FN are computed using number of different thresholds of the region growing algorithm and the ROC curve was generated between the $normalised\ sensitivity$ and $normalised\ (1-specificity)$. According to the ROC curve in Figure 3.11 the area under the curve was 0.9579. The optimal operating point on the ROC curve was selected such that the segmentation algorithm gives a higher sensitivity at the expense of FPs. In medical diagnosis, higher specificity is always important than the sensitivity as diagnosis tests should avoid FPs to prevent the treatment of non-affected individuals. However, in our case, the positively detected individuals will be monitored for other clinical symptoms for some period before diagnosing or treating for a disease. Therefore, detecting more positives is more important than avoiding FPs. Therefore, the point gave the normalised sensitivity 0.8992 and normalised specificity 0.9337 was chosen as the optimal cut off threshold.
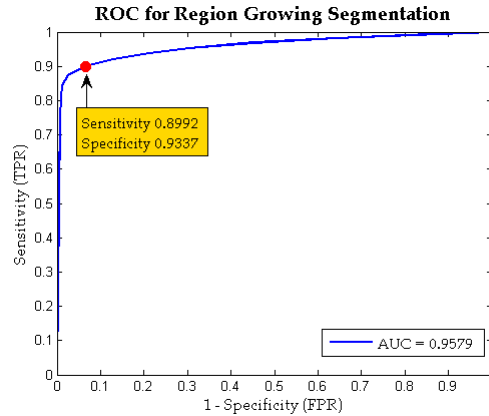
Figure 3.11: Receiver Operating Characteristic curve for the Region Growing Segmentation

Using the TP, FP, TN and FN calculated with the optimal cut off threshold we computed the MR, PRE, SEN, SPE, FPR, Accuracy (ACC), false positives per image window(FPPW) and f-score of the region growing segmentation of the test data set (Table 3.4).

Table 3.4: Validation results of the airway region growing segmentation algorithm for all the subjects

MR	PRE	SEN	SPE	FPR	ACC	FPPW	f-score
0.1864	0.9631	0.8136	0.9998	0.0002	0.9987	0.0172	0.8820

According to the validation results of the segmentation algorithm, it can be seen that the accuracy of the segmentation is 0.9987. However, human airway on an axial laryngeal CT image slice is a small area with respect to the total image size (usually 3% of the total image size). Therefore our data always has more negatives than positives. Due to this fact, obtaining a high accuracy in the segmentation step is trivial for this data.

As the high accuracy achieved, is uninformative in our case, the precision and sensitivity becomes more important in deciding on the performance of the algorithm. Validation on our test data set showed a 0.9631 precision and 0.8136 sensitivity. Furthermore, our algorithm works with a very low FPR of 0.0002. The MR of our method was 0.1864 and we only had 0.0172 FPPW. These higher values imply the accurate segmentation of human

airway in axial laryngeal CT scans.

However, it is also clear that images of some subjects appear as an exception to the normal validation results. As an example, observing the images of Subject 19 shows that the airway size of this subject is much smaller than average (Figure 3.12). Due to this reason there may not be enough texture information extracted from the region to detect the airway. So, the number of TPs detected will become low and number of FPs will be high, thus decreasing the sensitivity and precision values. Therefore, it is understood that the detection window size used in this experiment was not suitable for patients with smaller airway sizes like Subject 19.

Similarly, the airway size of Subject 13 is larger than the average. When a subject has larger airways a 96×96 pixels detection window cannot hold the whole airway inside (Figure 3.13). Therefore those detections were eliminated during the FPs removal step as partial detections. This explains the lower number of TPs in this subject.

Furthermore, the appearances of pharynx or laryngeal ventricles in some image slices (Subject 6) have caused the classifier to perform incorrectly (Figure 3.14). Most of the time, these additional darker holes were incorrectly detected as the airway and the true airway was missed by the algorithm. This lowers the sensitivity of our algorithm.
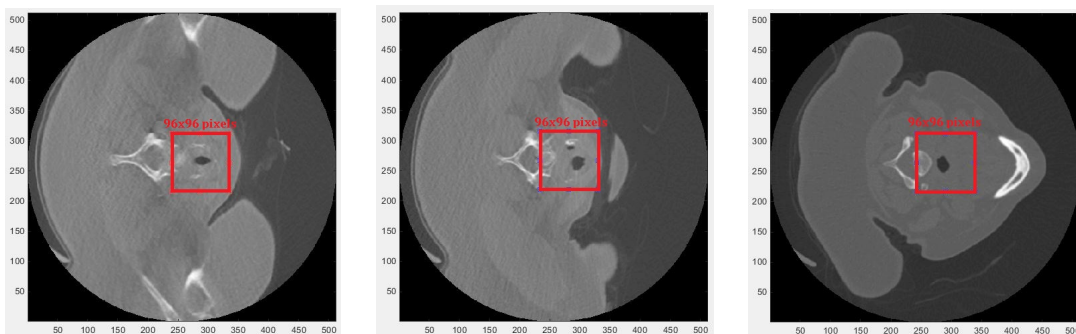
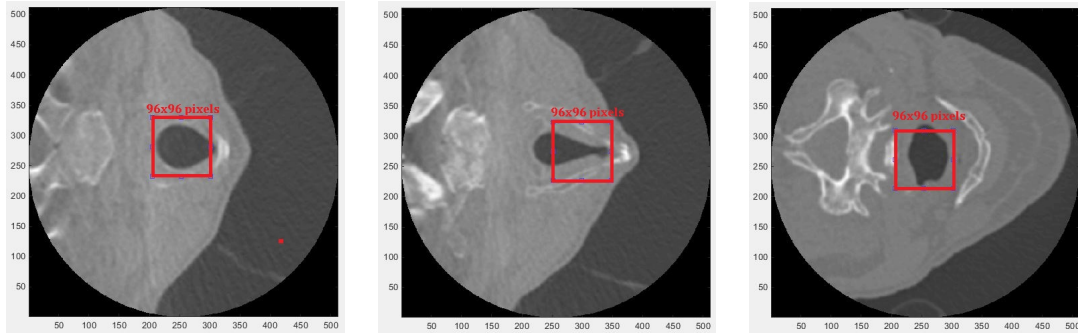*Figure 3.12: Subject 19 airway images: smaller airways*
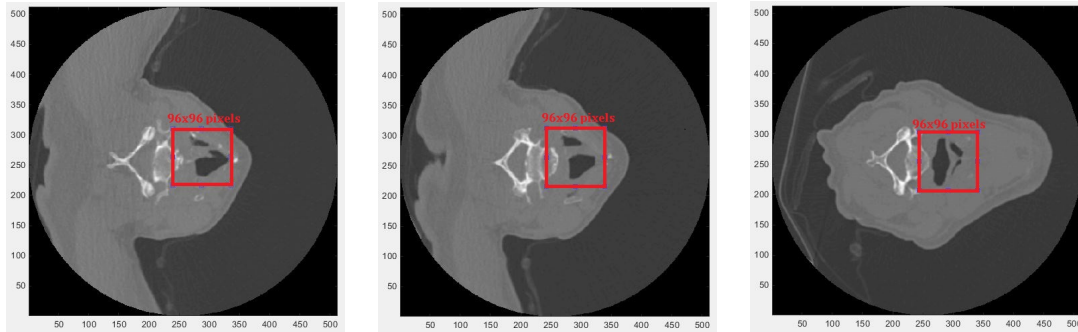
Figure 3.13: Subject 13 airway images: larger airways

Figure 3.14: Subject 6 airway images: appearance of similar darker areas

Validation of time series rima glottidis area curves

Finally, the detected rima glottidis area of each volume were used to generate the time series curves of 4D data of each patient. The ground truth time series rima glottidis area data for all the subjects were provided by the expert radiologists. The system generated time series curves were validated against this ground truth and few time series similarity indices were calculated to evaluate the performance of the whole system. Average correlation coefficient, average mean squared error, average Euclidean distance and average dynamic time warping distance between the computed and ground truth time series data for all the subjects in the test data-set were calculated and summarised in Table 3.5.

Table 3.5: Average time series similarity indices calculated for all the subjects

Dataset	Average correlation coefficient	Average mean squared error (mm^4)	Average Euclidean distance (mm^2)	Average dynamic time warping distance (mm^2)
All subjects	0.507	0.080	1.788	6.066
Best subjects (70%)	0.797	0.044	1.342	4.707

When observing the results of each subject of our test data-set, 70% of the subjects were found to be very highly correlated with their ground truth area with an average correlation coefficient of 0.797. However, 30% of subjects were found to be uncorrelated with correlation coefficients less than 0.3. Removing these low performing subjects from the data-set improved the average mean squared error to 0.044, average Euclidean distance to 1.342 and average dynamic time warping distance to 4.707.

This low performing data was found to be low quality images with noise and blurring. As an example, scans of Subject 5 were very noisy and blurred and in some of them an

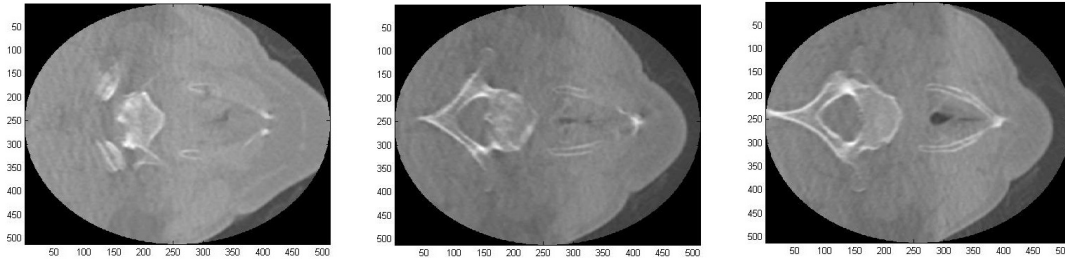airway was not detected at all (Figure 3.15).

Figure 3.15: Subject 5 airway images: noisy, blurred airway region

In some of that image volumes, the rima glottidis was appeared as two parts, which is called the "hourglass deformity" phenomenon. This produces the hourglass shape in rima glottidis at the time of closure of vocal folds. The reason could be due to a pathological condition on a vocal fold or blurring due to motion of the folds. During the manual segmentation, radiologists compute the two areas separately and add them together. However, during our experiment the algorithm only picks one part and gives a false area calculation (Figure 3.16). Furthermore, the appearances of laryngeal ventricles in some image slices (in Subjects 2, 4, 6, 9 and 17) resulted in failure. Most of the time, these additional darker holes were incorrectly detected as the airway (Figure 3.14).
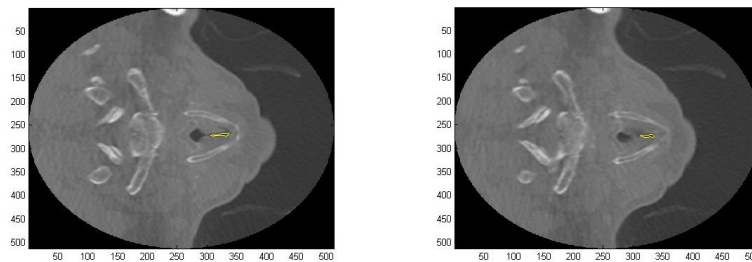
Figure 3.16: Hourglass deformity of rima glottidis in subject 2: the segmentation algorithm has picked only one region

Some subjects with below or above average vocal folds sizes and different airway shapes have also shown lower correlation values. When vocal folds are larger than normal (Subject 13), a 96×96 pixels detection window cannot hold the whole rima glottidis inside. Therefore, the rima glottidis images of those subjects were eliminated during the FPs removal step as partial detections (Figure 3.13).

Furthermore, as explained earlier in this section our SVM classifier does not work well when subjects have very smaller or misshaped vocal folds (Subject 19). Therefore, these subjects (Subject 19) shows a very low correlation value with the ground truth.

3.7   Discussion

In this chapter, a machine learning model based on texture features was proposed to segment the human vocal folds from volumetric laryngeal CT scans. The hypothesis suggested by Perju et al. [80] in 2014, shows an increase in glottic area of early PD patients due to inefficiently activated muscles and cartilages in the vocal area of larynx. When using this feature to early detect PD, the radiologists have to segment the vocal folds area manually to analyse the vocal dynamics. However, owing to the large volume of data available with the modern CT technology, the manual segmentation has become extremely time consuming, difficult and often giving poor reproducible results. Our aim was to provide a fully automated segmentation algorithm to compute the rima glottidis area, which can assist radiologists and doctors to predict PD. Furthermore, analysing the vocal folds movement is a crucial step when diagnosing many other voice-related pathologies. Therefore, this system has the potential of being used in those diagnosing procedures as well.

In this work, an automated algorithm was developed using SVMs to segment the human airway from axial CT scans. Texture features based on gray-level difference statistics in CT images were used as the features to segment the human airway area. An automated 2D region growing was then used to delineate the area inside the detected airway. The rima glottidis plane is the narrowest part in a human laryngeal cavity. Using this anatomical knowledge about human vocal folds, the rima glottidis plane was identified as the plane with the minimum airway area. Repeating these steps to extract the rima glottidis area of each volume in the 4D laryngeal data, the time series rima glottidis area curves were generated for all the subjects. These time series curves can be used to analyse the vocal dynamics of the patients and predict vocal abnormalities. Furthermore, this diagnostic information can be combined with the patient's clinical records to detect the

underline cause of the vocal abnormality early, such as PD, MS, laryngeal pathologies, etc.

The machine learning model combined with the post-processing module for FPs and redundancies removal proposed in this chapter, gave excellent results in classifying airway windows from 2D laryngeal CT images with a MR of 0.1304 at 0.0633 FPPI. Validation with the ground truth data showed a higher precision and sensitivity of 0.9279 and 0.8696, respectively. However, the algorithm did not work as expected when the glottis area is smaller or larger than the average size. To overcome this problem, the detection window size has to be adjusted such that it matches with the different vocal fold sizes. Therefore, multiple detection windows of different sizes can be used as a future improvement to the algorithm.

The 2D region growing algorithm we suggested in this chapter gave us very accurate delineations of the airway area from the classified windows of the SVM model. Our method traced around 81% of the airway pixels with 0.0172 FPPW. Approximately 96% of the pixels we detected were actual airway pixels overlapped with the ground truth. Yet, the region growing algorithm was sensitive to the noise and blurring presented in the images and adversely affected by the appearance of laryngeal ventricles or similar air-filled regions on 2D image slices.

Computed areas of the rima glottidis of each volume in the 4D data-set were used to generate time series curves. These curves generated for all the subjects in the test data-set were compared with the ground truth temporal maps of rima glottidis area provided by the experts. Seventy percent of the subjects in the data-set showed excellent correlation with the ground truth giving an average correlation coefficient of 0.797, average mean squared error of 0.044 and average Euclidean distance of 4.707. However, the remaining 30% of subjects in the test data-set were not correlated well with the ground truth. Most of the images in this subset were blurred, misshaped, appeared with pharynx or laryngeal ventricles or showed the hourglass deformity, hence did not perform well with our segmentation methods.

One major problem in the classification and segmentation methods proposed in this chapter were the sensitivity to the noise and blurring. To improve this, a pre-processing technique can be adapted to correct the noisy, blurred images. Also, the appearance of pharynx or laryngeal ventricles which are also similar darker areas, made the algorithm incorrectly identify these areas as vocal folds thus giving incorrect area calculations. In addition to the intensity-based features, structural features based on shape and size could be incorporated to the feature set to avoid detecting these FPs.

Moreover, according to the human laryngeal anatomy anterior edges of the vocal folds are connected to the AC on the thyroid cartilage and the posterior edges are connected to the vocal process of the arytenoid cartilages (Figure 3.17). Due to this reason, axial CT images planes are often not exactly aligned in the vocal folds plane. Instead, the actual vocal plane is positioned with a slight angle to the axial plane. Therefore, by adapting the method to align the axial plane in the plane of true vocal folds before processing the data will give better results.
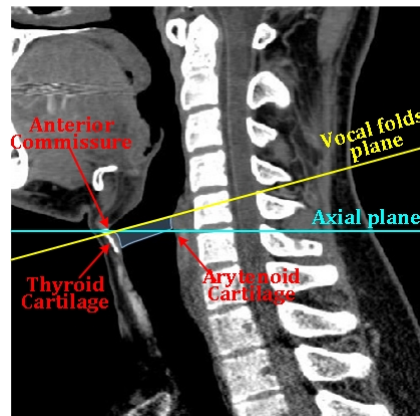
Figure 3.17: CT axial plane not aligned in the vocal folds plane

However, the results we got from the experiment reflect the fact that for about 70% of the patients, with well-correlated results with the ground truth, vocal fold plane detection which is a non-trivial process may not be required at all. Being the first work reported in this area, future research can be focused on detecting rima glottidis after adjusting the vocal plane and further improving the algorithm to achieve better results.

Chapter 4

Estimation of Vocal Folds Plane in 3D Laryngeal CT Images

According to the human anatomy, normal position of the vocal plane in a laryngeal CT image volume is not always aligned with the axial plane. This chapter provides an automatic computer vision-based method to estimate the vocal plane position in 3D laryngeal CT images. This work is based on the following publication.

S. Hewavitharanage, J. Gubbi, D. Thyagarajan, K. Lau and M. Palaniswami, "Estimation of vocal fold plane in 3D CT images for diagnosis of vocal fold abnormalities," 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, 2015, pp. 3105-3108.

4.1 Introduction

Vocal folds are two smooth bands of muscles located in the larynx just above the trachea (Figure 4.1). Humans produce voice by vibrating the vocal folds using the air coming from the lungs. The phonatory process occurs when air is expelled from the lungs through the glottis, creating a pressure drop across the larynx. Normally, the vocal folds remain closed until this laryngeal pressure is sufficient enough to move them apart. When they are opened, air escapes through the glottis reducing the pressure enough for

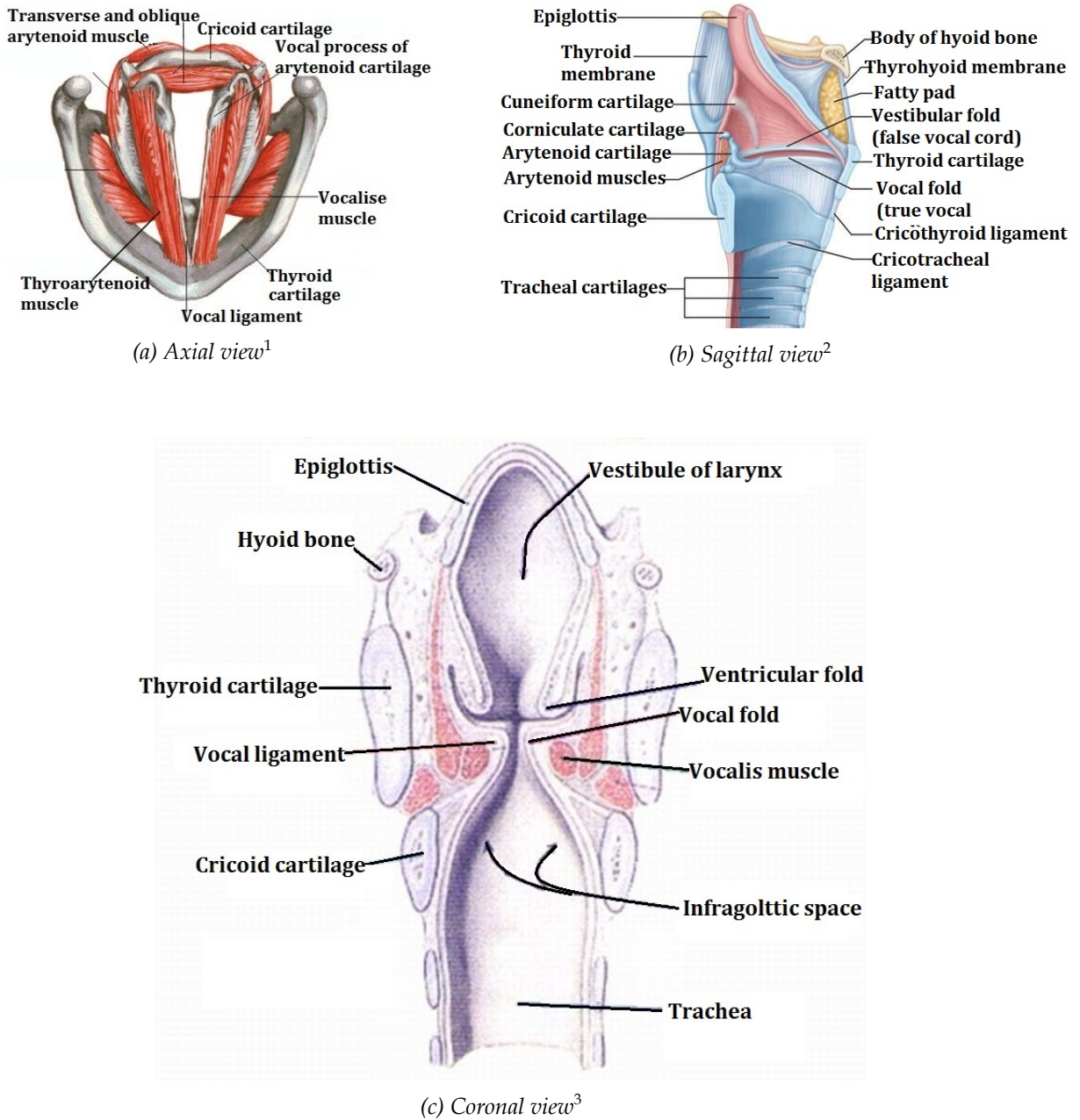the muscle tension to draw the folds back together again.

(a) Axial view[1]

(b) Sagittal view[2]

(c) Coronal view[3]

Figure 4.1: Different cross sectional views of human larynx

This abduction and adduction of vocal folds are controlled by the muscles connected

---

[1]https://www.studyblue.com/notes/note/n/ch-4-part-3/deck/5578458
[2]https://www.slideshare.net/carolinetokarski1/chapter-22-respiratory-system-1
[3]http://thesingingvoice.com/about/vocal-anatomy

to thyroid cartilage, cricothyroid cartilages and arytenoid cartilages. The lateral cricoarytenoid muscles contract and rotate the cricoarytenoid joint and pull the vocal process of arytenoid cartilages to bring the vocal folds together and close the rima glottidis (the gap between the vocal folds). Conversely, the contraction of posterior cricoarytenoid muscles rotate the cricoarytenoid joint and moves the vocal folds apart to open the rima glottidis (Figure 4.2).

(a) Full abduction   (b) Full adduction
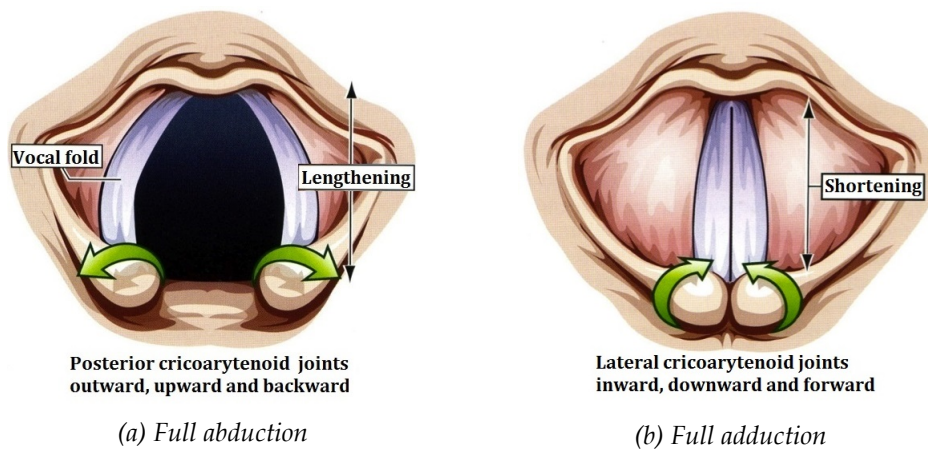
Figure 4.2: Movement of vocal folds and connecting cartilages during speech process[4]

When vocal muscles are misused or excessively used, they can be strained or damaged and vocal folds abnormalities may occur. Furthermore, vocal folds can be damaged and the connecting cartilages and muscles can be affected due to the effect of other illnesses like PD, MS, MG, strokes or tumours.

Previous studies carried out using laryngoendoscopy, laryngostroboscopy and laryngeal electromyography of PD patients [12], [13] found that those patients have an abnormal phase closure and abnormal laryngeal muscle activity. Furthermore, in 2014, Perju et al. [80] showed a significant difference in the laryngeal movement in patients with early PD. They used CT images of a group of PD patients and demonstrated that the inter-arytenoid distance is reduced due to inefficient vocalis and cricoarytenoid muscle activity in early PD.

Using the hypothesis above, a successful early diagnosis process can be developed

---

[4]https://www.studyblue.com/notes/note/n/commsd-4213-study-guide-2012-13-aulgar/deck/9719298

for PD. However, for this method to analyse the laryngeal muscle movement should exist. According to the studies done by Perju et al. [80], distances between the laryngeal cartilages can be used as a measure to analyse the laryngeal dynamics. However, the locations of the joints where vocal muscles and these cartilages meet should be segmented properly.

Laryngoendoscopy, laryngostroboscopy and laryngeal electromyography have been used for many years to explore the vocal folds of patients with vocal disorders [12], [13]. However, these methods only allow surface visualisation of the vocal folds. In most of the recent literature, CT technology has been used to detect the pathologies in vocal tract [5], [14], [15]. CT is a non-invasive, convenient 3D imaging technology and most modern CT scanners allow imaging of the entire larynx in a single breath-hold, thus avoiding the distortion caused by the respiratory motion or articulatory motion. Therefore, a better quantitative analysis can be achieved using CT images.

The joints where thyroid and arytenoid cartilages connect to vocal folds are located on the vocal folds plane. Most of the time, the vocal folds plane is not aligned in the normal axial plane of a CT image volume. So, to segment these joints first we have to find the vocal folds plane. The anterior edge of the vocal folds is connected to just below the posterior side of the superior AC. The posterior edges of the vocal folds are connected to the vocal process of arytenoid cartilages (Figure 1.7). Furthermore, the anatomical position of the vocal folds is affected by the patient position,motion, phonation and breathing during scanning. However, the thyroid and arytenoid cartilages cannot be visualised clearly on a neck CT image. Therefore, localising the actual vocal fold plane is a very complex task.

Various papers have been published to segment the airway and vocal tract in laryngeal CT images [5], [14]. However, no known research had been conducted to automate the vocal fold plane extraction process. Therefore, radiologists have to manually adjust the vocal fold plane parallel to the axial plane to measure the laryngeal parameters accurately [80]. However, due to the large amount of data involved in modern CT imaging, manual intervention is a tedious and time-consuming task. Therefore, we aim to de-

velop an automated algorithm to estimate the vocal fold plane using computer vision techniques.

In this work, we propose a method to automatically estimate the vocal fold plane using vertebral column and AC localisation. Using gray-level thresholding, connected component analysis, unsupervised k-means clustering and rule-based segmentation, MSP of the human larynx was estimated and the vertebral column and AC were localised. Then using known anatomical information the plane of the vocal folds was approximated. Being the first work reported in this area, we hope that this system will significantly reduce the work-load for the radiologists.

4.2 Data Selection and Generation

4.2.1 Data Selection

Ground truth data of AC were provided only for 35 subjects in the Parkinson-MONASH data-set. Therefore, 4D image data from those 35 subjects were used for validation in this study. Dataset consisted of 11 PD patients, 10 were controls and 14 uncategorised subjects, where 13 of them were females and 22 were males. All images were 4D laryngeal images.

4.2.2 Ground Truth Generation

Ground truth values for the location of AC was provided by the radiologists for this data set. However, manual annotation data for vocal plane positions was not provided. Therefore, performance validation of the vocal plane estimations was done by the radiologists by visual inspection.

4.3   Our Method

Extracting the vocal folds plane is not a simple straight forward process due to the different orientation of a patient during scanning and the different gray-level variations in a laryngeal CT image. Although locating the AC location and the vocal process of arytenoid cartilages will give the exact position of the plane, it is extremely hard to segment these cartilages in a laryngeal CT image. Usually these cartilages have gray-level variation similar to neighbouring tissues and may be hard to accurately locate even by an expert.

As can be seen in Figure 4.3, the vocal folds are usually located opposite the C5 cervical vertebra. As there is no standard for the starting position of the scanning of the neck, it is hard to recognise C5 vertebra on a sagittal plane of a CT volume. However, one can locate the posterior aspect of the AC in a laryngeal MSP due to the shape and the gray-level variation around that neighbourhood
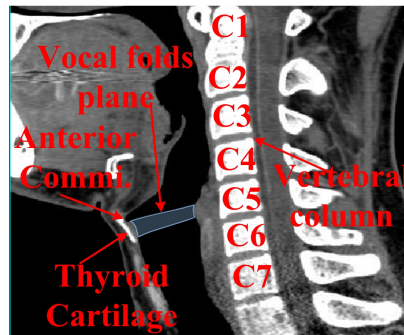
Figure 4.3: Anterior commissure, vocal folds plane and cervical vertebrae as seen on a laryngeal mid-sagittal plane

However, there is no such way to segment the vocal process of the arytenoid cartilages. These are the hardest parts to segment as these cartilages are extremely difficult to recognise even for a human observer. Therefore, we assumed there was not any coronal tilt in the images in the data-set and planned to follow an alternative way to get the estimated location of the vocal plane. For this, first we performed an initial investigation on the data-set and it showed that the vocal folds plane always makes approximately a $90°$ angle with the posterior edge of the vertebral column (Figure 4.4).
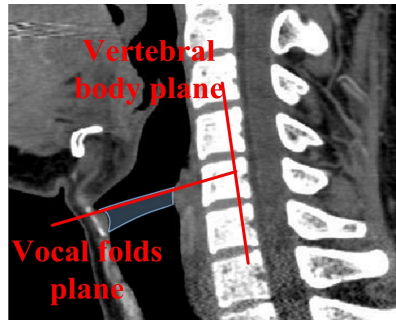
Figure 4.4: Vocal folds plane makes approximately $90°$ angle with the posterior edge of the vertebral column

Therefore, using this information and the assumption above, a complete automated system was developed to estimate the vocal folds plane location in a laryngeal CT image volume. The flowchart of the proposed framework is explained in Figure 4.5.

The 3D laryngeal CT image volume was processed using gray-level thresholding and connected component analysis to extract the MSP of the image volume. Then, thresholding and connected component analysis was combined with a rule-based segmentation algorithm to extract the anterior edge of the airway on the MSP and the edge profile was analysed using the known anatomical features to localise the AC. Again, the MSP was pre-processed with thresholding and connected component analysis and a k-means unsupervised classifier was used on the pre-processed MSP to segment the vertebral column. Finally, the plane of the vocal folds was approximated based on the locations of the AC and the vertebral column.

4.3.1 Thresholding

During the process of this algorithm gray-level thresholding was used many times to segment the regions of interest from the original images. Thresholding is the simplest processing technique which creates a binarised version of a gray-level image. This method uses a threshold intensity value T and normally replaces all the pixels in the image with intensity less than T with black pixels and pixels with intensity greater than T with white pixels or vice versa. The threshold value can be chosen according to the requirement of the problem and the image characteristics.
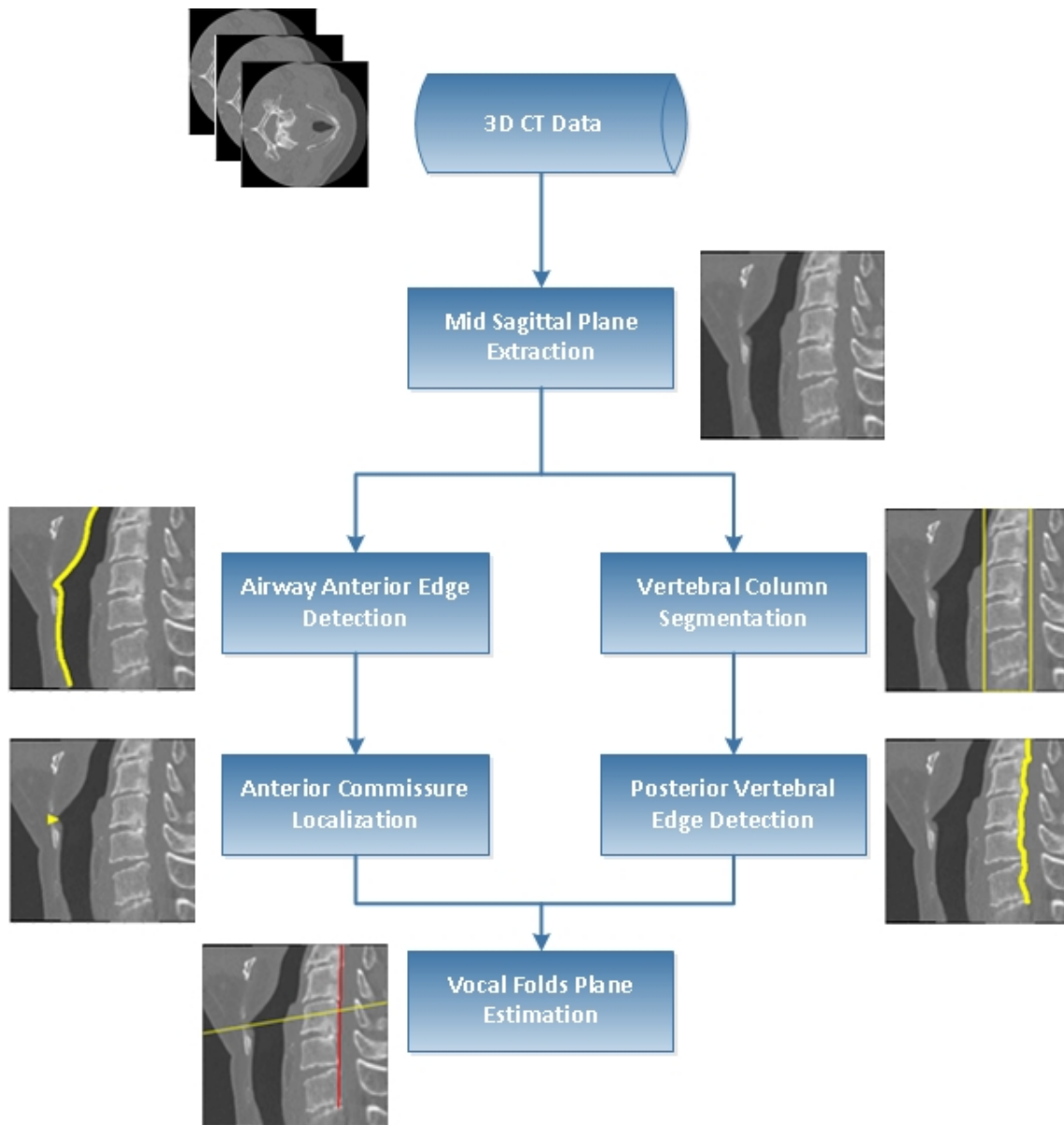
Figure 4.5: Vocal folds plane estimation framework

Based on the information manipulated by the algorithm, a few types of thresholding methods exist. Histogram-based methods [123] analyse the image histogram and find an optimal threshold which divides the histogram into two classes. Clustering methods [124] cluster the image into background and foreground. This is the most popular thresholding method. Entropy-based methods [125] use the entropy value to threshold the image. Spatial methods [126] use higher order probability density distribution functions to divide the image. Some thresholding methods use adaptive thresholding techniques [127], which change the threshold value used each time according to the local

image features. In this experiment we used a cluster-based thresholding technique to extract the background area, air-filled regions and bones from the laryngeal CT images.

To define a threshold value to eliminate the background regions from the CT images, all the background pixels of the original CT image were analysed and the minimum intensity value was taken as the threshold T_{BG} (Algorithm 2). For the data-set used in this experiment, the T_{BG} value was -2048HU.

Then we calculated two other threshold values for air-filled regions (T_D) and bones (T_B). For that a popular cluster-based thresholding technique, Otsu's method [121] was used (Algorithm 3). Otsu's method assumes the image only contains background and foreground and finds an optimal threshold which an image can be divided into those two classes such that the intra-class variance in both classes is minimal.

Executing the Algorithm 3 for air-filled areas and bones with respective initial threshold T_S with -1000 HU and $+700$ HU gave -436HU for T_D 482HU for T_B (Section 3.5.3).

4.3.2   Mid-Sagittal Plane Extraction

The human body is approximately symmetrical along the vertical plane through the mid line of the body, which is called the MSP. This plane divides the body into left and right halves passing through the mid-line structures like navel, spine and the airway [128]. There are number of publications in literature for extracting the MSP of human brain [129], [130], [131].

Ardekani et al. [129] suggested a method which used the cross correlation between two sides of the brain as the similarity measure. The plane which shows the maximum correlation was chosen as the MSP. In 2002, Prima [130] et al. proposed a similar method for 3D brain MRI images. They calculated the similarity measures in the two sides of the brain and identified the similar areas using a block matching procedure. Then the MSP was extracted as the plane that gives the maximum symmetry between those areas. This result was used to correct the tilt in head images. In 2011, Ruppert et al. [132] used the

correlation of the edge-detected MRI image and its flipped copy across the prospective plane as a similarity measure. They reported the maximum accuracy compared to the existing methods at that time.

Yuan et al. [133] used a template head model and some known landmarks to register a head CT image. Then by analysing the landmarks and known medical information the MSP was defined. This algorithm showed a better accuracy compared to the symmetry analysis methods. Schwing et al. [134] also used a landmark detection method to extract the MSP.

However, there are no known publications for extracting the MSP of the other parts of the body. As airway is located along the mid-line of the body, the MSP should pass through the centreline of the airway. The proposed method in this chapter (Figure 4.6) has used this information to extract the neck MSP from CT images.
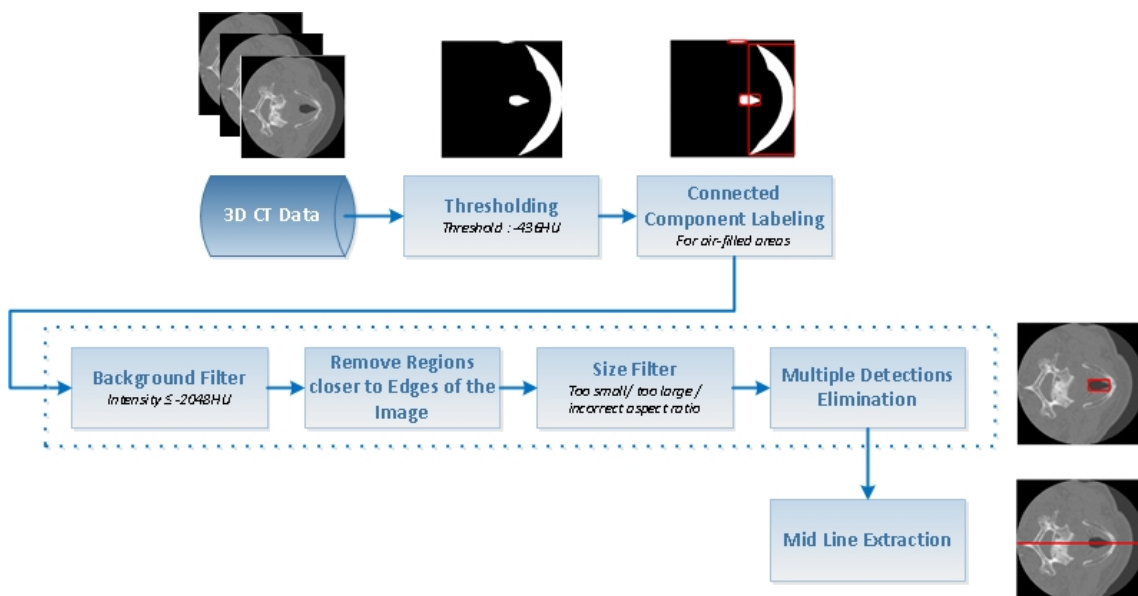
Figure 4.6: Mid-sagittal plane detection steps

The algorithm used the normal axial slices in a CT volume. First each axial slice was thresholded using a threshold T_D to divide the image slice into two clusters, where air-filled darker regions are the foreground. Then 8-connected neighbourhood connected component analysis was performed to extract the connected darker regions in the image.

Connected component analysis is a computer vision algorithm which labels all the connected pixels in a binary image based on a given heuristic. The image is scanned in the $8-$ or $4-$connected neighbourhood manner (Figure 3.5) and once a pixel in the connected region was found, all the connected pixels follow the given rule are also labelled before proceeding to the next pixel. This process continues until all the pixels in the foreground region are labelled.

Performing the connected component analysis on the thresholded axial image gave multiple air-filled regions. Therefore, to eliminate the other regions and select the required airway region a rule-based segmentation algorithm was employed (Equation 4.1). The rules were defined according to the known features of a human airway on an axial plane.

R_1: *Region width > Region height*
R_2: *Region width < Image width/2*
R_3: *Region height < Image height/2*
R_4: *Region width > 5% of the image width*
R_5: *Region height > 5% of the image height*
R_6: *Gray level of pixels in region neighbourhood > T_{BG}*
R_7: *Gray level of pixels in region > T_{BG}*

$$Region = \begin{cases} airway, & R_1, R_2, R_3, R_4, R_5, R_6, R_7 = TRUE \\ \\ non-airway, & otherwise \end{cases} \qquad (4.1)$$

Owing to the anatomical constraints of the human larynx, width of the glottis area is always larger than the height (Figure 4.7a). Furthermore, we only need to localise a number of airway windows to extract the symmetry axis. Too large or too small regions, regions with background pixels and adjacent to the background can be safely removed (Figure 4.7c), assuming that the airway is always surrounded by body pixels and both width and height are over 5% of the image width and height. Similarly, images with multiple regions (Figure 4.7b) detected also were removed from the outcome of the above
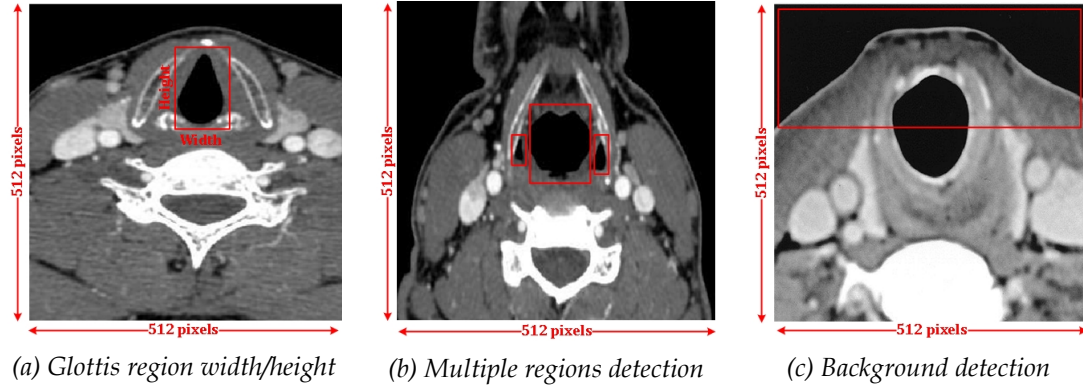
rule-based segmentation algorithm.

(a) Glottis region width/height   *(b) Multiple regions detection*   *(c) Background detection*

Figure 4.7: Rule-based segmentation output

Then the vertical bisector of the segmented airway region was taken as the axis of symmetry for the respective image slice and the mean of that of all the image slices in the CT volume was selected as the MSP of the volume.

4.3.3   Airway Anterior Edge Detection

Posterior aspect of the AC is located on the anterior edge of the airway on the MSP. Therefore, to find the location of the AC, the airway anterior edge should be detected on the mid-sagittal image calculated from the previous step (Figure 4.8).
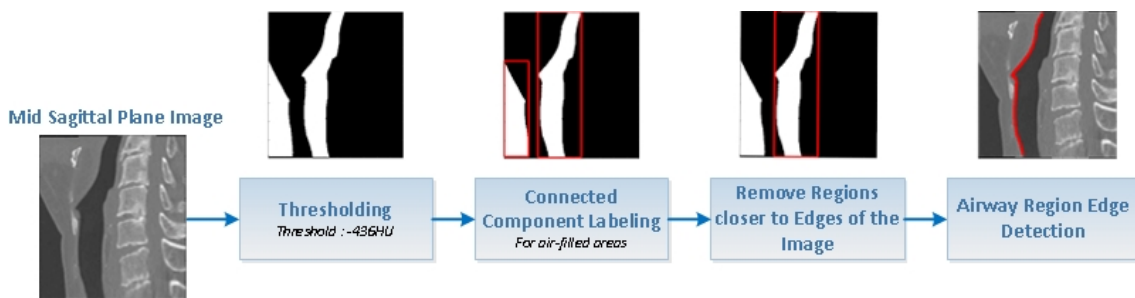
Figure 4.8: Airway anterior edge detection steps

Mid-sagittal image was thresholded using T_D to make a binary image with the air-filled regions as foreground. Then, 8-connected neighbourhood connected component analysis was done to label each connected air-filled region. Normally, a MSP image has

multiple air-filled region. However, only the airway region is surrounded by body pixels. All the other regions includes background pixels (Figure 4.9).
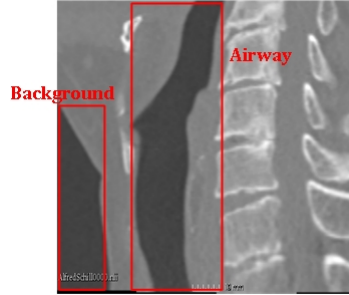
Figure 4.9: Airway and background region detections on mid-sagittal plane: airway is surrounded by body pixels

Therefore, the rule-based segmentation algorithm showed in Equation 4.2 was applied to eliminate the false detections and to select the airway.

$\mathbf{R_8}$: *Gray level of pixels in region neighbourhood* $> T_{BG}$

$$Region = \begin{cases} airway, & R_8 = TRUE \\ \\ non-airway, & otherwise \end{cases} \tag{4.2}$$

Then Moore-neighbour tracing algorithm [135] was employed to extract the posterior edge of the airway. Moore-neighbour tracing algorithm finds the boundary of an object in a 2D binary image. Moore-neighbourhood is the 8 pixels that share an edge of the central pixel in a 2D image. This is similarly known as 8-connected neighbourhood and shown from pixels $P1, P2, P3, P4, P5, P6, P7$ and $P8$ in Figure 4.10.
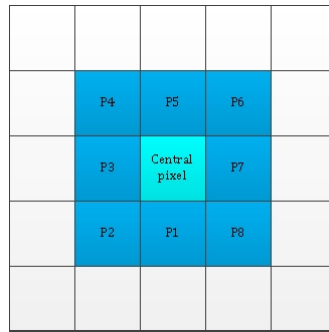
Figure 4.10: Moore neighbourhood pixels

In this algorithm, the binary image of the airway region was taken and starting from the bottom left corner, the image was scanned until an airway pixel was encountered. This pixel was used as the starting point. Once an airway pixel was found, the 8 pixels neighbourhood of that pixel was scanned in a clockwise direction and when another airway pixel was found it was labelled as a boundary pixel. The process was repeated until it visited the starting pixel in the same manner for the second time (Jacob's stopping criteria [135]). Using the resulting boundary pixel coordinates extracted, the boundary profile of the airway can be plotted as in Figure 4.11.

(a) Boundary coordinates profile

(b) Axial slice number profile
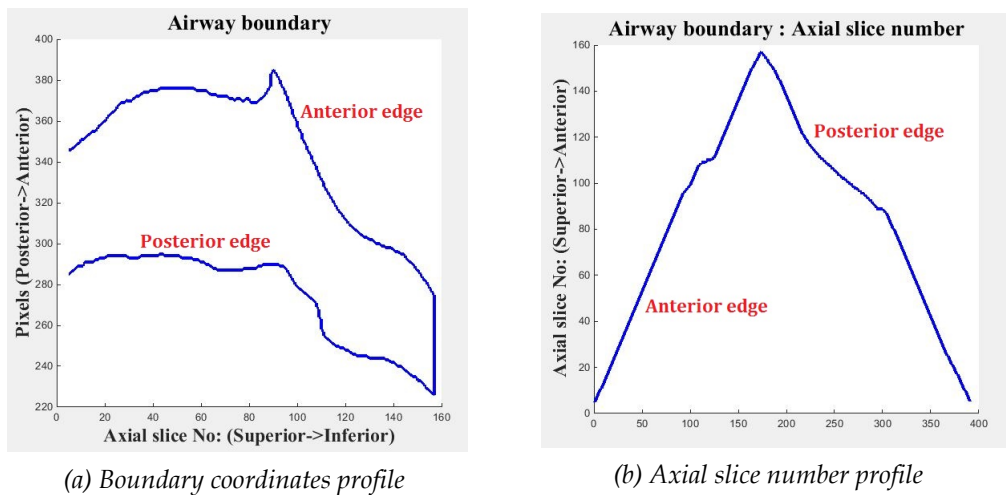
Figure 4.11: Airway boundary coordinates and axial slice index

Therefore, by filtering the boundary coordinates up to the peak in the axial slice number profile, the anterior edge coordinates can be retrieved.

4.3.4   Anterior Commissure Localisation

AC is located in the posterior side of the thyroid cartilage and it is the furthest point on the airway wall from the posterior border of the neck. According to the human anatomy, moving cranially from AC, this distance decreases significantly as shown in Figure 4.12. Therefore taking the peak location of the posterior border of the neck to the airway anterior edge distance profile, which has the subsequent maximum fall, should give the location of the AC. A sample distance profile is shown in Figure 4.12.
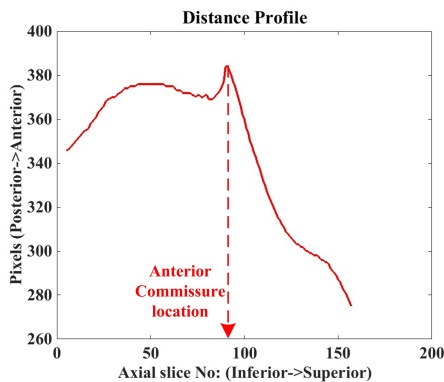
(a) Distance from the posterior border of the image to the anterior edge of the airway

(b) Anterior commissure location on the airway boundary

Figure 4.12: Distance profile of the anterior airway boundary

However, due to the posture differences in some patients this approach was not successful. Therefore, another known anatomical feature of the neck was used to filter the distance profile. When scanning from inferior to superior, airway diameter is approximately constant along the trachea but significantly decreases after the vocal folds as it shares the space with the pharynx (Figure 4.13). Taking this anatomical information into account, the distance profile was truncated after the location where airway diameter shows over 30% drop before computing the AC location. The threshold 30% was selected heuristically for the data set used.
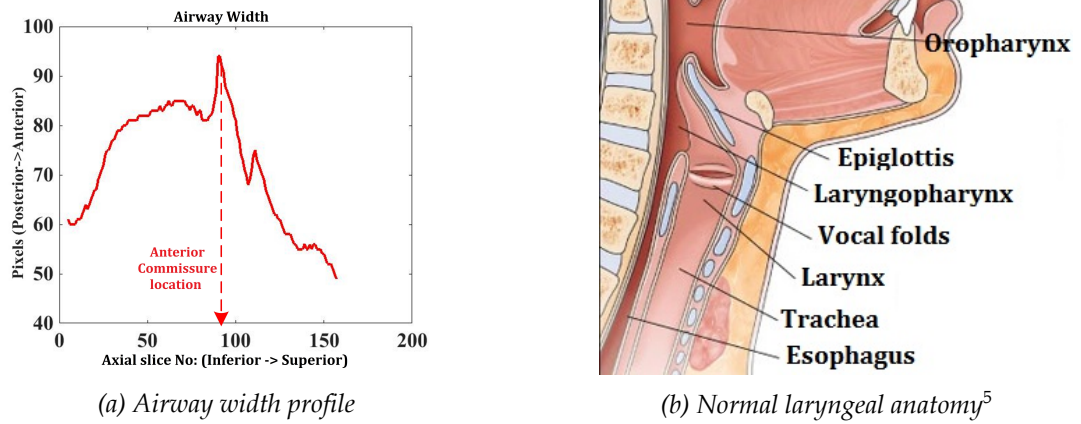
---

[5]http://orl.nc/pathologies-du-cou/pathologie-des-cordes-vocales/

(a) Airway width profile

(b) Normal laryngeal anatomy[5]

Figure 4.13: Changing of airway width with the axial anatomy

4.3.5 Vertebral Column Segmentation and Posterior Edge Detection

Vertebral bones are the brightest parts of a laryngeal CT image. Therefore, vertebral column can be visualised clearly on a MSP of a laryngeal CT volume. To compute a parallel plane to the posterior edge of the vertebral column, first the vertebral body should be segmented. The steps used to detect the vertebral column posterior edge is shown in Figure 4.14.
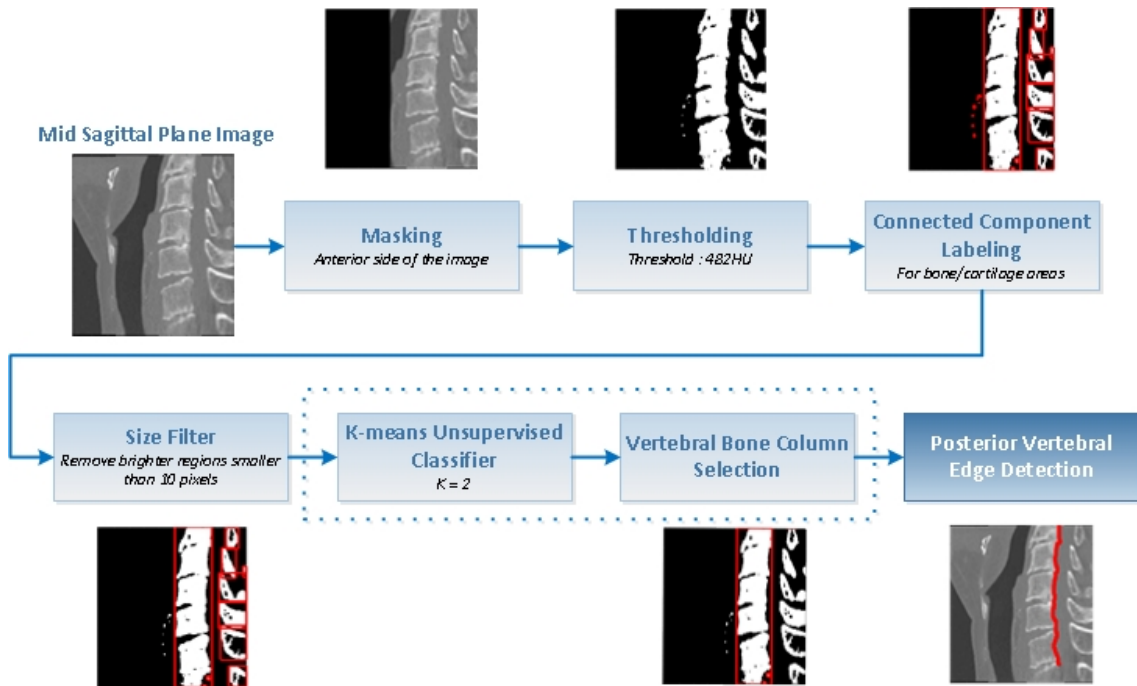
Figure 4.14: Vertebral column posterior edge detection steps

As vertebral bones are located on the posterior side of the airway, anterior side of the airway does not need to be processed during this step. Therefore, the anterior side of the MSP image was masked up to the posterior border of the airway. Then the resulting image was thresholded by T_B to localise brighter areas. The 8-connected neighbourhood connected component analysis was executed to find the connected brighter regions. This resulted multiple regions of bones and cartilages on the MSP image (Figure 4.15).
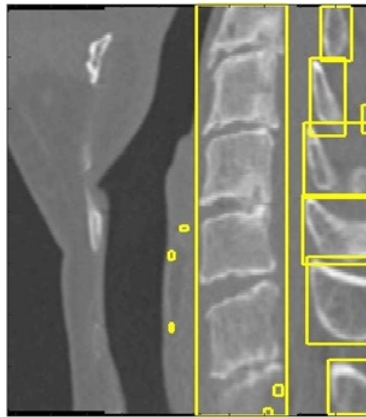

Figure 4.15: Brighter areas labelling on mid-sagittal plane

The main regions detected from the brighter areas labelling are the vertebral body, spinous process and the small parts of cricoid cartilage. As the vertebral body bones are relatively larger regions, a size filter was applied to this result to eliminate areas that have a pixel count less than 10. The threshold of 10 pixels were chosen heuristically using the anatomical knowledge. This resulted vertebral body and spinous process localisation.

These remaining regions were classified into different classes using a k-means unsupervised clustering. k-means is the simplest form of clustering used to classify n set of data to k number of clusters, where each data point belongs to the closest cluster. The objective is to define the k cluster centroids, one for each cluster, such that the square of the distances from the data points to the cluster centroid are minimum. The number of clusters depend on the data.

When x_i is n number of data points where $i = 1, 2, 3, ..., n$ which should be classified to k number of clusters, c_i is data points belongs to each cluster i and μ_i is the centroids of

each cluster, the k-means algorithm solves the objective function 4.3.

$$\arg \min_c \sum_{i=1}^{k} \sum_{j=1}^{c_i} d(x, \mu_i)^2 \tag{4.3}$$

$d(x, \mu_i)$ is the Euclidean distance between data point x and the centroid μ_i.

The steps to solve the locations of each centroid are shown in Algorithm 5.

Algorithm 5 : k-means Algorithm

STEP 1 : Centroids are placed as far as possible from each other

STEP 2 : Each data point is labelled to the nearest centroid based on the Euclidean distance from the data point to the centroids

STEP 3 : The new centroids are calculated as the mean cluster distance of all data points associated in that cluster

STEP 4 : Repeat STEP $2-3$ until there is no change in the centroid positions from one iteration to the next

---

As can be seen on the sagittal CT images, the vertebral body and spinous process have clearly different positions, distances from the posterior edge of the image to the geometric centres of the regions were used as the feature for the classifier. Furthermore, as the vertebral body and the spinous process are the main recognisable regions on the resulted filtered images, k was set to 2 in this experiment. From the output, the mean distance for each class was calculated and the class with maximum distance was taken as the vertebral body region.

Then the posterior edge of the vertebral body column was extracted using the same Moore-neighbour tracing algorithm [135], explained in the Section 4.3.3.

4.3.6 Vocal Plane Estimation

To extract the vocal folds plane, we used the empirical data evidence we obtained from analysing our data-set. It showed that the vocal folds are orthogonal to the posterior edge of the vertebral body column on MSP. We assume there was no any coronal tilt in the images. Therefore, first a parallel line to the posterior edge of the vertebral column on mid

sagittal plane was computed. Then the orthogonal line to that parallel line through the AC was extracted. From that we can estimate the angle, the vocal plane makes with the axial plane on mid sagittal plane. Assuming no coronal deformation of the data volume, now we can adjust the axial planes to align with the vocal folds plane.
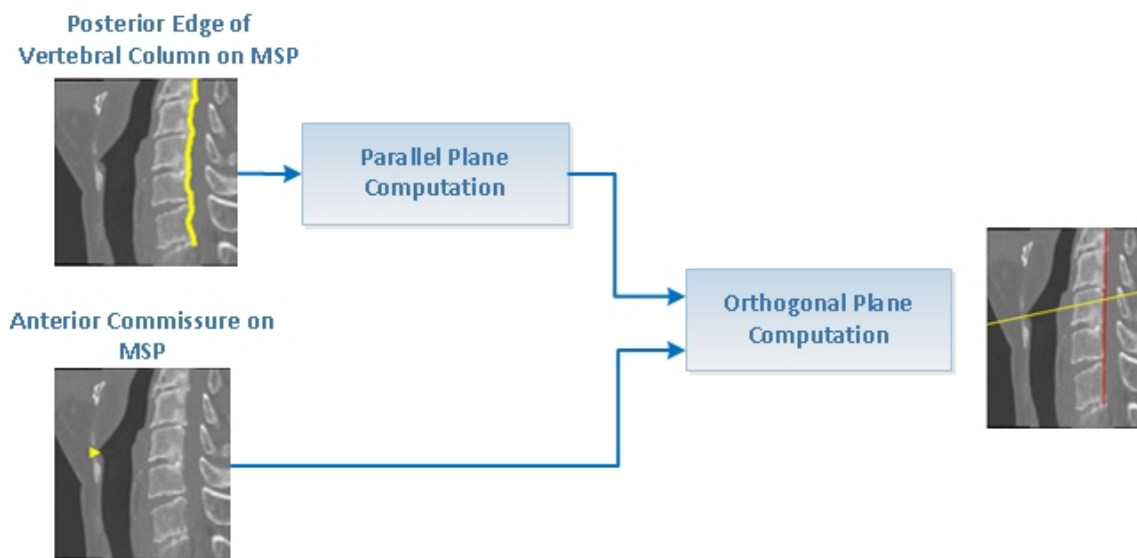
The flow chart for this step is shown in Figure 4.16.

Figure 4.16: Vocal fold plane estimation process

4.4 Results and Performance Validation

4.4.1 Results

The estimated vocal plane positions only for frame 1 of all subjects in the data-set are shown in Figure 4.17.
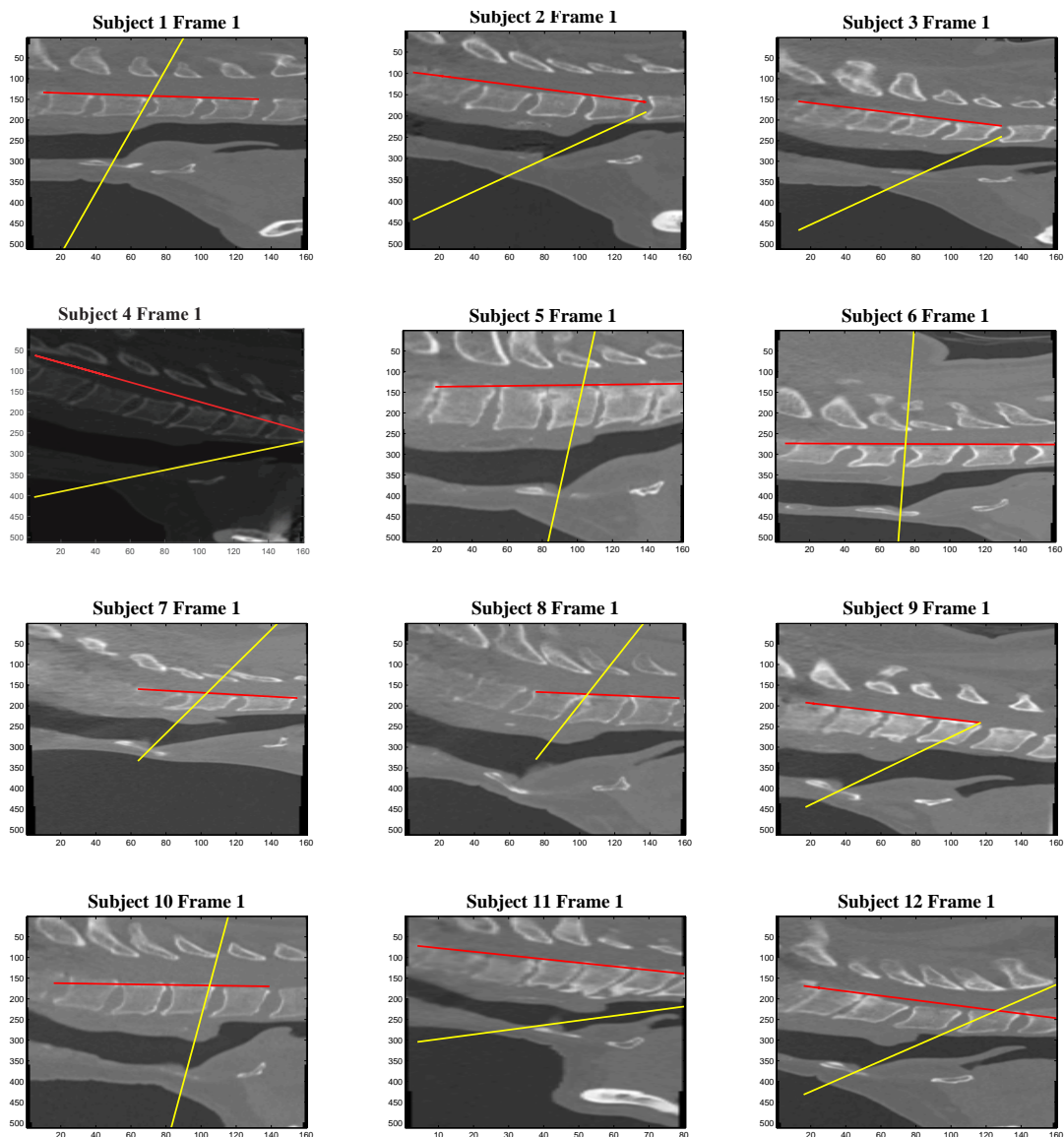
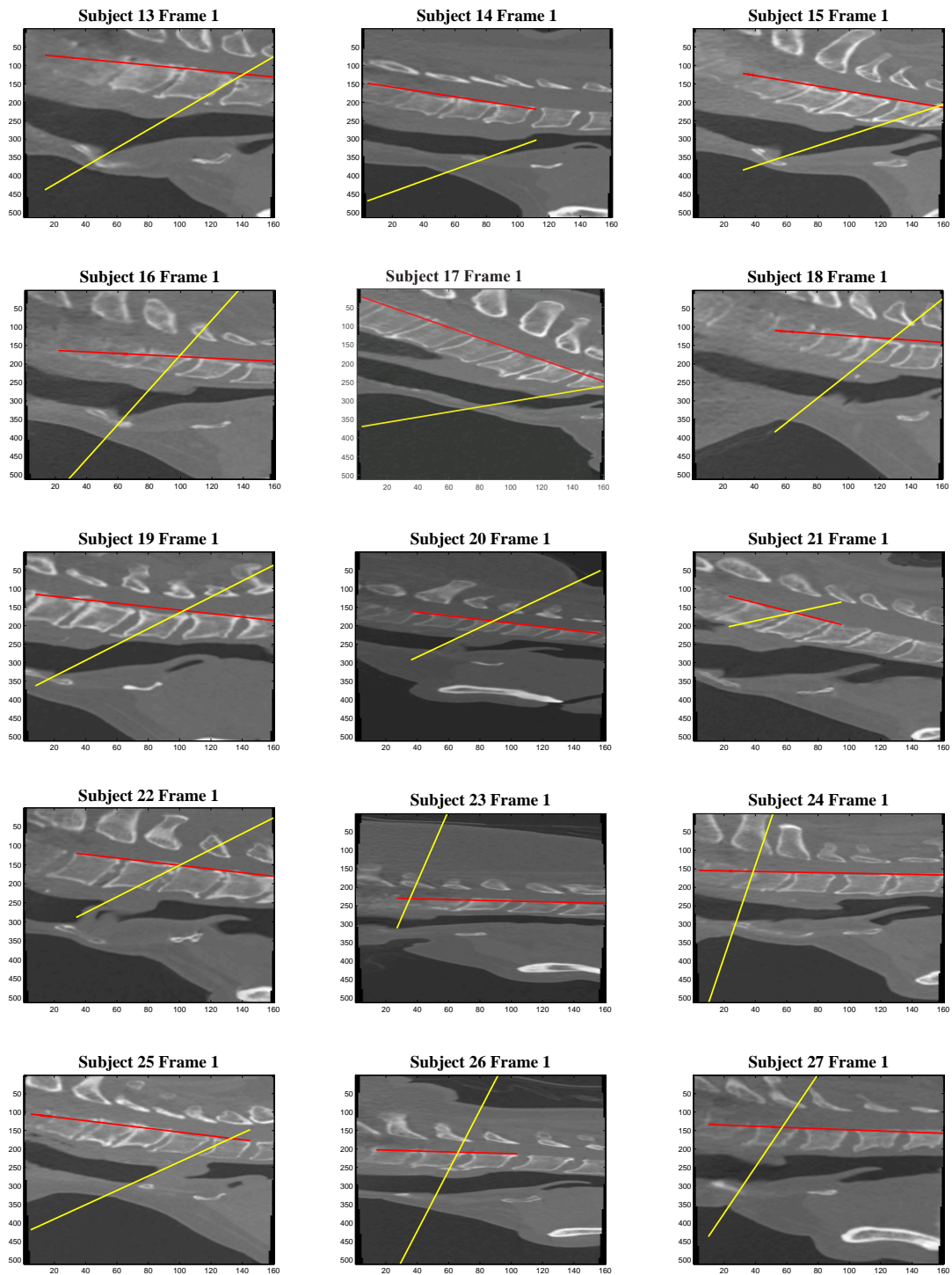*Figure 4.17: Estimated vocal plane locations of frame 1 of all subjects in the data-set*

Figure 4.17: (Cont.) Estimated vocal plane locations of frame 1 of all subjects in the data-set
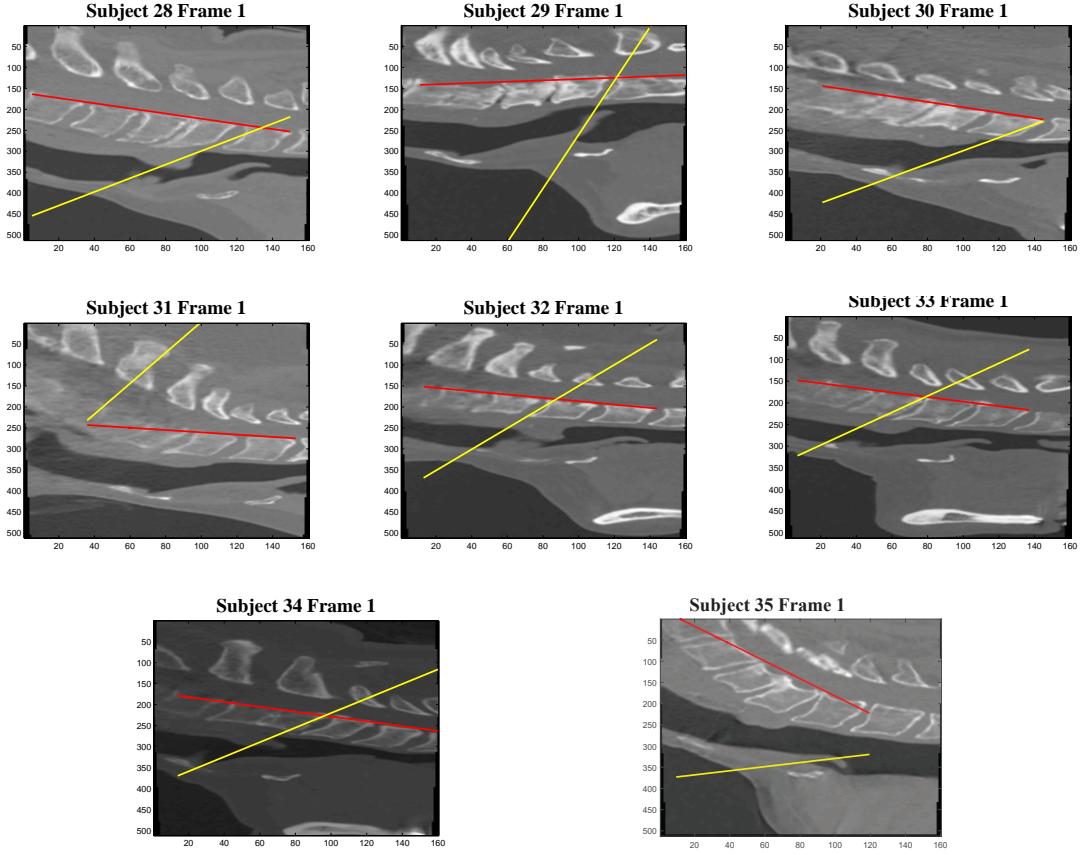
Figure 4.17: (Cont.) Estimated vocal plane locations of frame 1 of all subjects in the data-set

4.4.2 Performance Validation

Performance of the mid-sagittal plane detection

Ground truth MSPs were obtained from the ground truth coordinates of the AC locations provided by the radiologists. The error in MSP localisation was calculated as the absolute difference between the selected MSP slice index and the ground truth MSP slice index for each volume. When detected MSP index and ground truth MSP index of volume i are S_i and S_{Gi} the calculation of error in image volume i, e_i was done as in Equation 4.4.

$$e_i = |S_i - S_{Gi}| \tag{4.4}$$

Even though the voxel dimensions of the subjects in our data-set were different, the number of sagittal planes in a data volume were fixed for all the subjects, i.e. 512 slices. Therefore, we used the slice index for our calculations to normalise the effect of size vari-

ations between the subjects.

The number of *SUCCESS* or *FAILURE* detections were decided by applying different thresholds t. When the error, e_i, is not greater than t the detection was recorded as a *SUCCESS*, otherwise a *FAILURE*. Cumulative error index (CEI) was calculated for each threshold using Equation 4.5.

$$CEI_t = \frac{N_{e_i \leq t}}{n} \tag{4.5}$$

In Equation 4.5, $N_{e_i \leq t}$ is the number of image volumes that error e_i is not greater than threshold t. n is the total number of image volumes tested.

The cumulative error distribution curve was drawn using the calculated CEI_t for all thresholds (Figure 4.18).
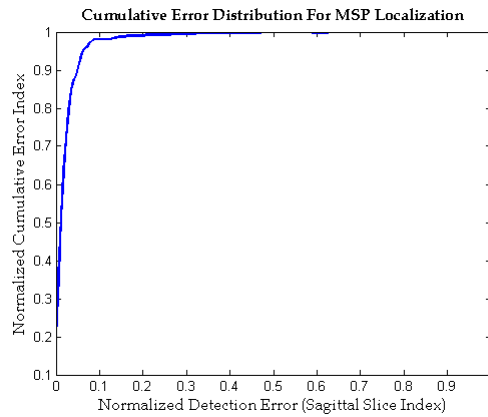

Figure 4.18: Cumulative error distribution curve for mid-sagittal plane detetction

According to the CED curve in Figure 4.18, approximately 80% of image volumes have MSP error of less than 0.03, which is roughly equal to ± 5 sagittal slices or $1 - 2.5\ mm$ in distance. Only a 2% of image volumes showed more than 10% of error. We computed the minimum (MIN), maximum(MAX), median, mean and standard deviation (STD) of the error values of the data-set. Furthermore, the precision and sensitivity of the data-set performance were calculated at the threshold error rate of 0.03 according to Equations 4.6 and 4.7. This threshold error (t) was decided considering the optimal performance point on cumulative error distribution curve.

$$Sensitivity = \frac{N_{success}}{N_{success} + N_{failure} + N_{missed}} \qquad (4.6)$$

$$Precision = \frac{N_{success}}{N_{success} + N_{failure}} \qquad (4.7)$$

In Equations 4.6 and 4.7, $N_{success}$ is the number of successfully detected volumes, $N_{failure}$ is the number of failed volumes and N_{missed} is the number of volumes that were unable to detect. Ground truth markings were provided for the total number of image volumes tested, which was also equal to $N_{success} + N_{failure} + N_{missed}$. Total number of detections were $N_{success} + N_{failure}$. The results are shown in Table 4.1.

Table 4.1: Results of the mid-sagittal plane detection method performance for all the subjects

MIN	MAX	Median	Mean	STD	Sensitivity	Precision
0	162	2	4.0136	8.0259	0.7845	0.7943

Performance of the anterior commissure localisation

To validate the performance of the AC localisation method, detected locations were compared with the manually marked values. Before starting the validation, both the system generated AC coordinates and the manual values were converted in to the respective anatomical coordinates (axial, sagittal and coronal values) using the voxel dimensions in the original NifTI headers of the data. The errors were calculated as the Euclidean distance between the detected point and the ground truth point. Errors calculated for all the volumes of the subjects were averaged and normalised to compute the mean normalised error of each subject. Normalisation was done using inter-cornu distance, which is an anatomically unique value for a subject, so that the various laryngeal sizes will not affect the results.

If Euclidean distance between the detected and manual values of AC of kth volume is d_k, 4D image data of the subject consists of n frames and inter-cornu distance of the subject is d_{icd}, the mean normalised distance error is given by Equation 4.8.

$$m_e = \frac{1}{n \cdot d_{icd}} \sum_{k=1}^{n} d_k \tag{4.8}$$

Then using different values to threshold the normalised error m_e, the $SUCCESS$ or $FAILURE$ number of detections were computed for the data-set. The CEI for each threshold, CEI_t was calculated using the number of $SUCCESS$ detections at error threshold t, $N_{m_e \leq t}$ and the total number of subjects N (Equation 4.9).

$$CEI_t = \frac{N_{m_e \leq t}}{N} \tag{4.9}$$

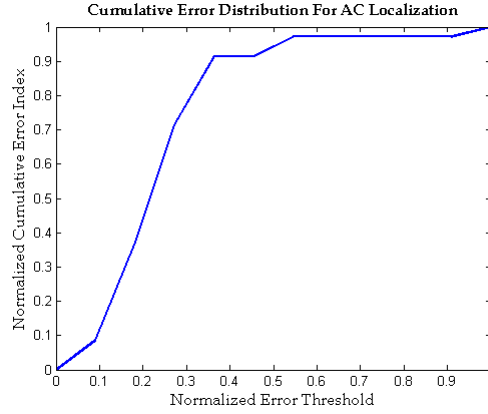The cumulative error distribution curve is shown in Figure 4.19.

Figure 4.19: Cumulative error distribution curve for anterior commissure detection

According to the CED curve in Figure 4.19, 91% of subjects have AC detection error of less than 0.4 of inter-cornu distance of the subject, which is roughly equal to $12 - 18\ mm$ in distance. Only a 3% of the subjects showed more than 0.6 of inter-cornu distance error. We computed the MIN, MAX, median, mean, STD of the error values of the data-set and precision (Equation 4.7) and sensitivity (Equation 4.6) at the error rate of 0.4.

The results are shown in Table 4.2.

Table 4.2: Results of the anterior commissure detection method performance for all the subjects

MIN	MAX	Median	Mean	STD	Sensitivity	Precision
0.0241	1.0916	0.2408	0.2603	0.1841	0.7065	0.8072

However, the AC is not a single pixel point on the human larynx. It is covered with a mucosal membrane lining and the normal thickness of the AC is about 1 mm [136]. Furthermore, human vocal folds are consisted of soft pilable tissue layers called epithelium, lamina propria and vocalis muscle. Therefore, adult human vocal folds have an average thickness of $3-5$ mm [4] as well. Moreover, manual markings of the AC that was used to validate the system performance were highly subjective and a sufficient level of tolerance should be considered when comparing. Considering all these facts an error tolerance of around 1 cm for the error distances between the system generated locations and the manually marked locations of the AC can be considered as reasonable.

Furthermore, some image frames of the data-set were very blurred and substantially distorted due to the movement of the vocal folds. It was hard to locate the AC in these images even for a manual observer. As an example, MSP images of Subject 1 are extremely distorted after frame 8 and the AC cannot be located (Figure 4.20).
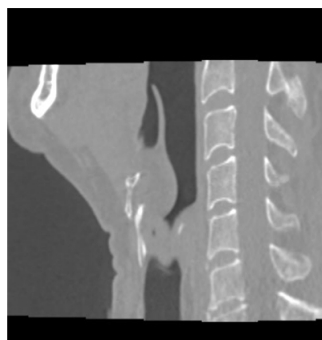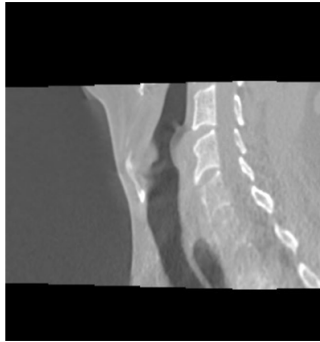

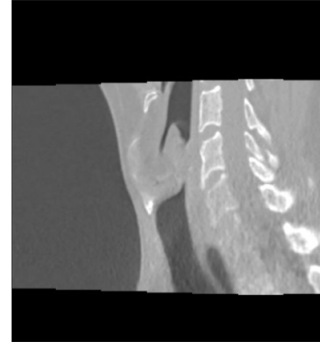*(a) Frame 9*   *(b) Frame 20*

Figure 4.20: Subject 1 mid-sagittal planes: very unclear anterior commissure

Similarly, Subject 7 has very unclear MSPs after frame 7. Therefore, AC was not de-

tected in these problematic frames (Figure 4.21). In some frames, the distortion was so high, even the MSP calculation step was not successful.
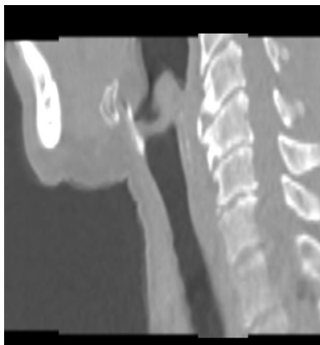

(a) Frame 9   *(b) Frame* 20

Figure 4.21: Subject 7 *mid-sagittal planes: very unclear anterior commissure*

However, in Subject 11, which only has 20 frames, the vocal folds appear as a joint structure due to the movement (Figure 4.22). In such frames, the AC location was not correctly detected.
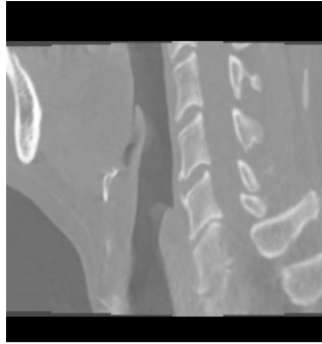

(a) Frame 14   *(b) Frame* 20

Figure 4.22: Subject 11 *mid-sagittal planes: joint vocal folds due to movement*

Moreover, due to the changes in the scan position, vocal folds were not presented in some subjects (Subject 21) at all. Other examples of some ungradable images are shown in Figure 4.23.

(a) Subject 21:*No vocal folds present on the mid-sagittal plane*

(b) Subject 21:*Very low resolution images*

Figure 4.23: Some subjects with ungradable mid-sagittal planes

Performance of vocal plane estimation method

There was no ground truth provided for validating the vocal folds plane position. Therefore, the estimated vocal folds planes from the system were visually assessed by the experts to find the accuracy of the proposed method and found to be concordant with the expectations.

However, we understand that visual assessment is highly subjective and not a standard validation method. But unfortunately, manual gold standard data do not exist and therefore coming to a conclusion regarding the performance of this method is difficult at the moment. However, it must be noted that finding the vocal folds plane on a CT image is very difficult even when it is done manually and outcome is always an approximation.

4.5   Discussion

In this chapter, a complete automated system to estimate the vocal folds plane from 3D laryngeal CT images was proposed. Based on the studies done by Perju et al. [80] in 2014, PD patients have a reduced inter-arytenoid distance due to the differences in vocalis muscles and surrounding cartilages. To use this hypothesis as a diagnostic feature to detect PD, medical professionals have to measure the laryngeal parameters such as inter-arytenoid distance and thyro-arytenoid distance. However, according to the human laryngeal anatomy the vocal folds are usually not aligned with the normal axial

plane of a CT volume. Therefore, measuring these laryngeal parameters is not a straight forward task. Radiologists have to manually align the CT volume in the plane of vocal folds before taking the measurements. Manually adjusting the CT data is always time consuming, owing to the large data-set available in modern CT imaging. Our aim is to provide an automated tool to estimate the vocal folds plane on a 3D laryngeal CT volume, which can assist radiologists in diagnostic procedures.

In this work, an automated algorithm was developed using AC localisation and the vertebral column segmentation. Computer vision techniques like gray-level thresholding, connected component analysis and anatomical feature-based rule-based segmenting were used to extract the MSP of a CT volume. Then the airway on the MSP was segmented and an edge detection algorithm was applied to detect the anterior edge of the airway. Next thresholding, connected component labelling and unsupervised k-means clustering were used to delineate the vertebral body column on the MSP. Resulting vertebral region was processed with an edge detection algorithm to extract the posterior edge of the vertebral column. Our investigation on the data-set showed the vocal folds plane is approximately orthogonal to the posterior edge of the vertebral column. Finally, using this anatomical information the plane of the vocal folds was approximated.

Testing of our MSP detection method on a subset of Parkinson-MONASH data-set reported that in 80% of images volumes the MSP was detected with less than $1 - 2.5$ mm error. The precision and accuracy of the method were 0.7943 and 0.7845. Our AC localisation system segmented the AC with an error of less than $12 - 18$ mm for 91% of the subjects tested. The precision and sensitivity of AC segmentation were 0.8072 and 0.7065.

However, these methods were sensitive to image quality, movement and posture differences. In future work, a pre-processing step to reduce the blurring and noise can be introduced to improve the results.

Furthermore, ground truth data was not available for the vocal folds plane and therefore a complete validation of the vocal folds estimation system could not be achieved. However, visual assessment showed a reasonably good approximation.

The exact plane of vocal folds passing through the AC on anterior side and vocal process of arytenoid cartilages on posterior side. However, arytenoid cartilages show very little gray-level variation compared to the surrounding tissue on a laryngeal CT image, thus segmentation of the structure is extremely difficult. Therefore, we used the anatomical knowledge about the data-set as an alternative way to find the vocal folds plane. However, if we could find a suitable way to delineate the arytenoid cartilages then more accurate results of estimating the vocal folds plane could be achieved.

However, it should be noted that selecting the vocal folds plane is always a hard task owing to the poor gray-level variations of these cartilages and the result is always an approximation. Being the first work in this area we report the state-of-the-art results of estimating vocal folds plane from 4D laryngeal CT images. Future work should be focused on segmenting the arytenoid cartilages to improve the accuracy.

Chapter 5

Conclusions

Voice disorders are a very common type of disease that can be seen in any age group. Apart from damage, tumours and other pathologies occur in vocal folds and cartilages, many neurological disorders like PD, MS and MG can cause various voice-related problems. Various modern medical imaging techniques like stroboscopy, endoscopy, CT and MRI are used to observe the human vocal folds when diagnosing voice disorders. Among them, CT imaging provides faster and better visualisation and hence mostly used in viewing the human larynx. However, clinicians have to manually adjust the vocal folds plane and segment the region before diagnosing as there are no known automated segmentation system to delineate the vocal folds region from a laryngeal CT volume.

This thesis proposed a fully automated segmentation tools to estimate vocal plane location and segment the vocal folds region. In this chapter, we sum up the contributions made by this project and address the limitations of the work. Finally, we will suggest some future research directions.

5.1 Summary of Contributions

In summary, this thesis proposed and validated a set of methods for automated measurement of the rima glottidis area and estimation of vocal folds plane position. A supervised SVM-based machine learning model and some methods based on computer vision were

proposed for segmentation of rima glottidis from 2D laryngeal axial images. The exact vocal plane position is slightly angled to the axial plane of the human larynx, therefore our next step was to develop a tool to estimate the vocal plane position using computer vision algorithms. A summary of these contributions is presented in the following paragraphs.

The first contribution of this research is a machine learning model for classifying of airway windows from 2D laryngeal axial CT images. Real volumetric CT images of larynx of PD patients and healthy subjects, selected from Movement Disorder Clinic of Monash Medical Centre, were used to develop and validate the proposed systems. Textural features were calculated on each sub-window of 2D axial image slice using GLDMs. A SVM classifier was trained with this feature set to recognise the patterns in 2D axial data and identify the airway windows. Final classification windows were post-processed to remove overlapping detections and to select a single window per image slice. The proposed model was evaluated on 20 subjects of the Parkinson-MONASH data set. The results show that our model achieves a high accuracy in classifying airway windows.

The second contribution of this research is the proposal of novel method for segmentation of the rima glottidis region and generation of rima glottidis area time series for the 4D laryngeal CT images. Anatomical information and intensity information were used to develop a novel computer vision algorithm to segment the rima glottidis area of each volume from the resulting prediction windows from our airway classifier. These delineations were incorporated with the voxel information of the patient to generate time series curves of the rima glottidis areas. The proposed algorithm was validated against the manual gold standard rima glottidis markings and time series curves and showed well-correlated results.

The third contribution of this research is a novel algorithm to segment the AC from volumetric laryngeal CT images. Using computer vision algorithms, intensity information and spatial information were analysed to detect the AC location. AC is the location where vocal folds are anteriorly connected to the laryngeal wall/thyroid cartilage. By detecting this we expect to find the exact position of the vocal plane. Our predicted AC

locations showed higher accuracy with the manual observer's markings.

Our fourth contribution of this research is a novel hypothesis-driven method to estimate the vocal folds plane position. By using known human anatomical information we approximated the vocal plane to a surface that is orthogonal to the vertebral column and passing through the AC. The resulting estimated vocal plane positions were visually validated by the experts and found to be concordant with the true vocal planes.

5.2   Limitations

Chapter 3 of this thesis provides an approach to automatically delineate the human vocal folds from a human laryngeal CT volume using a SVM. Due to the substantial gray-level difference of the rima glottidis region compared to the surrounding soft tissues on a 2D axial CT image, the texture features calculated on 2D axial CT images were used for classification. Automated 2D region growing algorithm was applied to the resulting rima glottidis windows of the SVM and the rima glottidis area was delineated.

The proposed rima glottidis segmentation system demonstrated excellent results in terms of area overlap with the ground truth data provided by the experts in the Monash Medical Centre. However, the developed system for rima glottidis segmentation was sensitive to noise present in the original CT images. We did not use any pre-processing method to remove blurring due to noise or movement from the images as de-blurring algorithms are normally time consuming and introduce artefacts in images which are not acceptable in medical image processing. These artefacts will hide real pathological details and interfere with the decisions.

Furthermore, investigating the causes for the differences in the system output and the ground truth area values revealed a significant problem in the automated process. Although axial planes were considered when calculating the rima glottidis area, due to the human laryngeal anatomy and the posture difference in different subjects during scanning, the vocal folds plane is usually not aligned with the normal axial plane of the CT

volume. Owing to this reason, the total rima glottidis area may not be included in the axial plane. During the manual process clinicians adjust the CT volume to align with the vocal plane before taking rima glottidis area measurements. Therefore, our automated algorithm also should incorporate a vocal plane selection step before applying the segmentation algorithm.

Our next work covered in this thesis was done to fulfil this requirement. Chapter 4 provides an automated approach to estimate the vocal folds plane from a laryngeal CT volume. However, finding the vocal folds plane from a laryngeal CT volume is a complex task even for an expert as landmarks define the position of the vocal plane are difficult to visualise on a CT image. Vocal folds are anteriorly connected to the AC of the thyroid cartilage. AC can be detected on a laryngeal mid-sagittal CT image due to the specific shape and the gray-level difference with the surrounding body pixels. However, the posterior ends of the vocal folds are connected to the vocal process of the two arytenoid cartilages. Arytenoid cartilages cannot be clearly visualised on a CT image as they have a very similar gray-level compared to the surrounding soft tissues. Therefore, an alternative method was used to decide the other end of the vocal plane using the known information about the data-set and human anatomy.

The initial observation on the data-set revealed that the vocal folds plane is approximately orthogonal to the vertebral bone column. Vertebral bones are the brightest part on a CT image, thus can be segmented easily using that intensity information. Then using the orthogonal feature, vocal folds plane was estimated as the plane passing through the AC making a $90°$ angle with the vertebral bones. However, this approximation was based on the assumption that the CT data has no coronal tilt.

The AC locations were compared with the manually marked location by the experts and found to be concordant. However, we did not have any manual gold standard data to validate the estimated vocal folds planes. These results were observed and confirmed by the experts only by visual observation.

Although we know that the visual observation cannot be accepted as a standard way

of validating the results, it should be understood that marking the vocal planes on these data is a complex task even for the manual observers. Due to not having proper landmarks that can be seen on a CT image, the manual process is also subject to errors. Furthermore, there may be situations that the data present with some coronal tilt due to the posture variations of the subjects. Our automated system does not perform well with that data.

5.3   Future research direction

Being the first work reported in this area, our system performed well in estimating the vocal folds plane and segmenting the rima glottidis area from laryngeal CT volumes. We hope these fully automated systems will be beneficial to the clinicians when measuring laryngeal parameters to diagnose voice disorders.

However, to use these systems as a complete diagnostic tool to detect voice disorders some improvement, additions and modification may be required as future work. This section briefly describes some potential future research problems which may come out of this work.

Validation with manual gold standard: First, manual gold standard data should be collected in order to validate the system generated output for the estimated positions of vocal fold planes. As manual markings are highly subjective at least markings of three different independent observers are required to improve validation. A graphical user interface (GUI) to obtain the user input for the vocal folds plane location of each CT volume was developed and explained in Appendix B of this thesis.

Pre-processing to reduce noise and blur: Blurring in images due to noise and patient or cartilages movement was another problem in our data. As abnormalities in vocal dynamics should be observed during speech, movement of the cartilages and vocal folds are unavoidable. Due to some neurological conditions like PD, patient movements are

also not easy to avoid. As a result, important details may be hidden and segmenting the cartilages and bones are not possible in such images. Most existing de-blurring techniques are very complex and time consuming. Furthermore, they cannot be adapted to medical imaging as they introduce unwanted artefacts to the data. Therefore, a better algorithm should be developed as a future improvement to reduce the blurring so that the sensitivity to the noise and movement of the system could be reduced.

Correction of the deformations: The coronal tilt in the laryngeal volume was not taken into account when developing our automated system for estimating the vocal folds plane. However, in real world situations the coronal direction tilt will present most of the time. Furthermore, the laryngeal volume may have other geometrical deformations like shear, translations and rotations. Any of these deformations were not taken into account when developing this system. Hence, a suitable registration technique with a reference data volume should be incorporated as a pre-processing step to improve the accuracy. As thyroid cartilage and vertebral bones are important landmarks on a laryngeal CT, first we can identify these landmarks and then use a registration mechanism based on them. Intensity-based registration could also be employed as the laryngeal CT volume-specific intensity pattern.

Detection of aryteoid cartilages: The most accurate way of locating the vocal folds plane is to detect the vocal process of the arytenoid cartilages rather than calculating the orthogonal plane to vertebral column. Yet, we found that segmenting the arytenoid cartilages from a laryngeal CT image is extremely difficult due to the low gray-level variation of cartilages with the surrounding soft tissues. However, these arytenoid cartilages have a specific pyramidal shape and texture which can be used to develop a new algorithm to segment them in future. This could be an important future development to this work.

Complete diagnosis tool to diagnose voice disorders: This thesis only provided two main components of a large diagnosis tool to detect voice disorders. As the first work reported in this field, we have developed the state-of-the-art methods. However, the

systems need some improvements and the potential of these systems to be used as sophisticated diagnostic tools are yet to be fully realised. As a major future research direction, we suggest a complete diagnosis tool for voice disorders as shown in Figure 5.1.
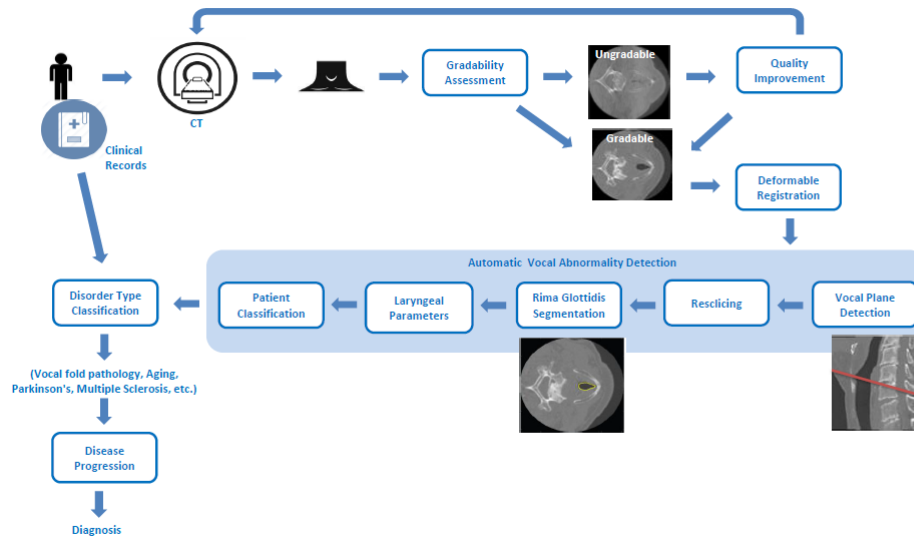

Figure 5.1: Block diagram of the complete voice disorders diagnosis workflow

After acquisition, human laryngeal CT images will first need to go through a quality assessment process to evaluate the gradability of the images. Ungradable images will then need to be pre-processed to enhance the quality. If the quality enhancement cannot achieve the required gradability, those patients have to be rescanned. Gradable images will first go through a deformable registration algorithm to align the image volumes with a reference standard. Our vocal plane detection method can be applied on these aligned image volumes to detect the position of the vocal plane. After that, the image volumes will be resliced in the angle of vocal plane, so that at the level of vocal folds axial slices represents the vocal plane. Then our rima glottidis segmentation algorithm can be used on these axial images to segment the vocal folds region. Additionally, an algorithm to segment thyroid cartilage and arytenoid cartilages also will be required at this stage.

Once the laryngeal cartilages and rima glottidis are segmented, laryngeal parameters like vocal folds area, inter-arytenoid distance, thyro-arytenoid distances can be calculated for each data volume. These parameters of each volume can be used in a separate

classifier to classify healthy and unhealthy subjects. Another classifier can be used with patient's clinical records and the results of the patient classifier as inputs, to detect the voice disorder type. However, subsequent disease progression analysis should be conducted in order to decide the exact cause of the voice disorder before appropriate clinical diagnosis.

Evaluation of generalisability: In our system, AC and vertebral bone segmentation, feature extraction, airway classification and rima glottidis segmentation were validated on a small data-set. Hence, the generalisability of our system is difficult to conclude. In order to use these system components in a complete diagnosis tool, it requires clinical testing on a substantially large data-set, to actually reduce radiologists' time and effort in manual analysis and improve the diagnostic accuracy.

Bibliography

[1] M. S. Ashutosh, R. Razvi, and G. S. Longia, "Clinical anatomy of the vocal cord," *People's Journal of Scientific Research*, vol. 1, pp. 35–38, 2008.

[2] J. N. Bhatia, O. P. Sachdeva, and A. K. Gupta, "Neglected foreign body in the larynx," *Indian Journal of Otolaryngology*, vol. 30, no. 3, pp. 138–138, 1978.

[3] U. Cinar, C. Yigit O Fau Vural, S. Vural C Fau Alkan, S. Alkan S Fau Kayaoglu, B. Kayaoglu S Fau Dadas, and B. Dadas, "Level of vocal folds as projected on the exterior thyroid cartilage," no. 0023-852X (Print), 2003.

[4] M. S. Hahn, B. A. Teply, M. M. Stevens, S. M. Zeitels, and R. Langer, "Collagen composite hydrogels for vocal fold lamina propria restoration," *Biomaterials*, vol. 27, no. 7, pp. 1104–1109, 2006.

[5] R. J. Holmes, D. J. Oates Jm Fau Phyland, A. J. Phyland Dj Fau Hughes, and A. J. Hughes, "Voice characteristics in the progression of parkinson's disease," *International Journal of Language and Communication Disorders*, vol. 35, no. 1368-2822 (Print), pp. 407–18, 2000.

[6] M. Liotti, D. Ramig Lo Fau Vogel, P. Vogel D Fau New, C. I. New P Fau Cook, R. J. Cook Ci Fau Ingham, J. C. Ingham Rj Fau Ingham, P. T. Ingham Jc Fau Fox, and P. T. Fox, "Hypophonia in parkinson's disease: neural correlates of voice treatment revealed by pet," *Neurology*, vol. 60, no. 1526-632X (Electronic), pp. 432–440, 2003.

[7] Y. Stelzig, V. Hochhaus W Fau Gall, A. Gall V Fau Henneberg, and A. Henneberg, "[laryngeal manifestations in patients with parkinson disease]," *Laryngorhinootologie*, vol. 78(10), no. 0935-8943 (Print), pp. 544–51, 1999.

[8] A. K. Ho, R. Bradshaw Jl Fau Iansek, and R. Iansek, "For better or worse: The effect of levodopa on speech in parkinson's disease," *Movement Disorders*, vol. 23, no. 1531-8257 (Electronic), pp. 574–80, 2008.

[9] C. Moreau, J.-L. Ozsancak C Fau Blatt, P. Blatt Jl Fau Derambure, A. Derambure P Fau Destee, L. Destee A Fau Defebvre, and L. Defebvre, "Oral festination in parkinson's disease: biomechanical analysis and correlation with festination and freezing of gait," *Movement Disorders*, vol. 22, no. 0885-3185 (Print), pp. 1503–6, 2007.

[10] S. Sapir, C. Ramig L Fau Fox, and C. Fox, "Speech and swallowing disorders in parkinson disease," *Current Opinion in Otolaryngology and Head and Neck Sugery*, vol. 16, no. 1531-6998 (Electronic), pp. 205–10, 2008.

[11] S. Sapir, J. L. Ramig Lo Fau Spielman, C. Spielman Jl Fau Fox, and C. Fox, "Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech," *Journal of Speech Language and Hearing Research*, vol. 53(1), no. 1558-9102 (Electronic), pp. 114–25, 2010.

[12] K. S. Perez, M. E. Ramig Lo Fau Smith, C. Smith Me Fau Dromey, and C. Dromey, "The parkinson larynx: tremor and videostroboscopic findings," *Journal of Voice*, vol. 10(4), no. 0892-1997 (Print), pp. 354–61, 1996.

[13] A. P. Zarzur, G. d. N. Duarte Is Fau Holanda, M. A. U. R. Holanda Gdo N Fau Martins, and M. A. Martins, "Laryngeal electromyography and acoustic voice analysis in parkinson's disease: a comparative study," *Brazilian Journal of Otorhinolaryngology*, vol. 76(1), no. 1808-8686 (Electronic), pp. 40–3, 2010.

[14] T. Ogawa, A. Enciso R Fau Memon, J. K. Memon A Fau Mah, G. T. Mah Jk Fau Clark, and G. T. Clark, "Evaluation of 3d airway imaging of obstructive sleep apnea with cone-beam computed tomography," *Studies in Health, Technology and Infomatics*, vol. 111, no. 0926-9630 (Print), pp. 365–8, 2005.

[15] I. Cheng, S. Nilufar, C. Flores-Mir, and A. Basu, "Airway segmentation and measurement in ct images," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, 2007, Conference Proceedings, pp. 795–799.

[16] C. L. Gordon, J. D. Webber Ce Fau Adachi, N. Adachi Jd Fau Christoforou, and N. Christoforou, "In vivo assessment of trabecular bone structure at the distal radius from high-resolution computed tomography images," *Physics in Medicine and Biology*, vol. 41, no. 0031-9155 (Print), pp. 495–508, 1996.

[17] H. R. Singleton and G. M. Pohost, "Automatic cardiac mr image segmentation using edge detection by tissue classification in pixel neighborhoods," *Journal of the Scociety of Magnetic Resonance in Medicine*, vol. 37, no. 0740-3194 (Print), pp. 418–24, 1997.

[18] M. Brewster, "An introduction to wavelets (charles k. chui)," *SIAM Review*, vol. 35, no. 2, pp. 312–313, 1993.

[19] J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-8, no. 6, pp. 679–698, 1986.

[20] L. S. Davis, "A survey of edge detection techniques," *Computer Graphics and Image Processing*, vol. 4, no. 3, pp. 248–270, 1975.

[21] A. RV and B. J, "Combining wavelet transform and hidden markov models for ecg segmentation," *EURASIP J. Appl. Signal Process.*, vol. 2007, no. 1, pp. 95–95, 2007.

[22] Q. Xujia, J. Jionghui, W. Weihong, and Z. Fan, "Canny operator based level set segmentation algorithm for medical images," in *Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on*, 2007, Conference Proceedings, pp. 892–895.

[23] R. Adams and L. Bischof, "Seeded region growing," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, no. 6, pp. 641–647, 1994.

[24] S. A. Hojjatoleslami and J. Kittler, "Region growing: a new approach," *Image Processing, IEEE Transactions on*, vol. 7, no. 7, pp. 1079–1084, 1998.

[25] D. Ji, K. Foong, and S. Ong, "A two-stage rule-constrained seedless region growing approach for mandibular body segmentation in mri," *International Journal of Computer Assisted Radiology and Surgery*, vol. 8, no. 5, pp. 723–732, 2013.

[26] L. Rusko, G. Bekes, and M. Fidrich, "Automatic segmentation of the liver from multi- and single-phase contrast-enhanced ct images," *Medical Image Analysis*, vol. 13, no. 6, pp. 871–882, 2009.

[27] E. M. van Rikxoort, B. de Hoop, S. van de Vorst, M. Prokop, and B. Van Ginneken, "Automatic segmentation of pulmonary segments from volumetric chest ct scans," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 4, pp. 621–630, 2009.

[28] K. Aloui and M. S. Naceur, "3d brain tumor segmentation using level-sets method and meshes simplification from volumetric mr images," *World Academy of Science, Engineering & Technology*, vol. 3, no. 33, p. 127, 2009.

[29] P. Lei, G. Lixu, and X. Jianrong, "Implementation of medical image segmentation in cuda," in *Information Technology and Applications in Biomedicine, 2008. ITAB 2008. International Conference on*, 2008, Conference Proceedings, pp. 82–85.

[30] V. Grau, M. Mewes Au Fau Alcaniz, R. Alcaniz M Fau Kikinis, S. K. Kikinis R Fau Warfield, and S. K. Warfield, "Improved watershed transform for medical image segmentation using prior information," *IEEE Trans Med Imaging*, vol. 23, no. 0278-0062 (Print), pp. 447–58, 2004.

[31] G. Hamarneh and X. Li, "Watershed segmentation using prior shape and appearance knowledge," *Image and Vision Computing*, vol. 27, no. 1-2, pp. 59–68, 2009.

[32] H. P. Ng, S. H. Ong, K. W. C. Foong, P. S. Goh, and W. L. Nowinski, "Medical image segmentation using k-means clustering and improved watershed algorithm," in *Image Analysis and Interpretation, 2006 IEEE Southwest Symposium on*, 2006, Conference Proceedings, pp. 61–65.

[33] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[34] H. El and J. M. Palomo, "Measuring the airway in 3 dimensions: a reliability and accuracy study," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 137, no. 1097-6752 (Electronic), 2010.

[35] T. Schlathoelter, C. Lorenz, I. C. Carlsen, S. Renisch, and T. Deschamps, "Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy," in *Proc SPIE Conf on Medical Imaging: Image Processing*, 2002, Conference Proceedings, pp. 103–113.

[36] K. Mori, J. Hasegawa, J. Toriwaki, H. Anno, and K. Katada, "Recognition of bronchus in three-dimensional x-ray ct images with applications to virtualized bronchoscopy system," in *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, vol. 3, 1996, Conference Proceedings, pp. 528–532 vol.3.

[37] T. Kitasaka, K. Mori, Y. Suenaga, J.-i. Hasegawa, and J.-i. Toriwaki, *A Method for Segmenting Bronchial Trees from 3D Chest X-ray CT Images*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2003, vol. 2879, book section 74, pp. 603–610.

[38] J. Tschirren, G. Hoffman Ea Fau McLennan, M. McLennan G Fau Sonka, and M. Sonka, "Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose ct scans," *IEEE Trans Med Imaging*, vol. 24, no. 0278-0062 (Print), 2005.

[39] B. van Ginneken, E. M. Baggerman W Fau van Rikxoort, and E. M. van Rikxoort, "Robust segmentation and anatomical labeling of the airway tree from thoracic ct scans," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2008*, vol. 5241, 2008, Conference Proceedings, pp. 219–226.

[40] H. Singh, M. Crawford, J. Curtin, and R. Zwiggelaar, *Automated 3D Segmentation of the Lung Airway Tree Using Gain-Based Region Growing Approach*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, vol. 3217, book section 118, pp. 975–982.

[41] M. Nakamura, T. Wada S Fau Miki, Y. Miki T Fau Shimada, Y. Shimada Y Fau Suda, G. Suda Y Fau Tamura, and G. Tamura, "Automated segmentation and morphometric analysis of the human airway tree from multidetector ct images," *The Journal of Physiological sciences*, vol. 58, no. 1880-6546 (Print), pp. 493–8, 2008.

[42] M. Sonka, P. Wonkyu, and E. A. Hoffman, "Rule-based detection of intrathoracic airway trees," *Medical Imaging, IEEE Transactions on*, vol. 15, no. 3, pp. 314–326, 1996.

[43] W. Park, M. Hoffman Ea Fau Sonka, and M. Sonka, "Segmentation of intrathoracic airway trees: a fuzzy logic approach," *IEEE Trans Med Imaging*, vol. 17, no. 0278-0062 (Print), pp. 489–97, 1998.

[44] L. Tan Kok, T. Tanaka, H. Nakamura, T. Shirahata, and H. Sugiura, "Segmentation of airway trees from multislice ct using fuzzy logic," in *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, 2009, Conference Proceedings, pp. 1614–1617.

[45] X. Ziyue, U. Bagci, B. Foster, and D. J. Mollura, "A hybrid multi-scale approach to automatic airway tree segmentation from ct scans," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, 2013, Conference Proceedings, pp. 1308–1311.

[46] D. Bartz, D. Mayer, J. Fischer, S. Ley, A. del Rio, S. Thust, C. P. Heussel, H. U. Kauczor, and W. Strasser, "Hybrid segmentation and exploration of the human lungs," in *Visualization, 2003. VIS 2003. IEEE*, 2003, Conference Proceedings, pp. 177–184.

[47] T. Bulow, C. Lorenz, and S. Renisch, "A general framework for tree segmentation and reconstruction from medical volume data," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2004*, ser. Lecture Notes in Computer Science, C. Barillot, D. Haynor, and P. Hellier, Eds., vol. 3216. Springer Berlin Heidelberg, 2004, Book Section, pp. 533–540.

[48] M. W. Graham, J. D. Gibbs, and W. E. Higgins, "Robust system for human airway-tree segmentation," in *SPIE Medical Imaging 2008: Image Processing*, vol. 6914, 2008, Conference Proceedings, pp. 69 141J–69 141J–18.

[49] R. A. Ochs, J. G. Goldin, F. Abtin, H. J. Kim, K. Brown, P. Batra, D. Roback, M. F. McNitt-Gray, and M. S. Brown, "Automated classification of lung bronchovascular anatomy in ct using adaboost," *Medical Image Analysis*, vol. 11, no. 3, pp. 315–324, 2007.

[50] P. Lo and M. de Bruijne, "Voxel classification based airway tree segmentation," in *SPIE Medical Imaging 2008: Image Processing*, vol. 6914, 2008, Conference Proceedings, pp. 69 141K–69 141K–12.

[51] P. Lo, J. J. H. Sporring J Fau Pedersen, M. Pedersen Jj Fau de Bruijne, and M. de Bruijne, "Airway tree extraction with locally optimal paths," in *MICCAI International*

Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 12, 2009, Conference Proceedings, pp. 51–8.

[52] P. Lo, J. Sporring, H. Ashraf, J. J. H. Pedersen, and M. de Bruijne, "Vessel-guided airway tree segmentation: A voxel classification approach," *Medical Image Analysis*, vol. 14, no. 4, pp. 527–538, 2010.

[53] C. Pisupati, "Geometric analysis of dynamic three-dimensional tree structures," PhD Thesis, The Johns Hopkins University, 1996.

[54] F. J. Preteux, C. I. Fetita, and P. Grenier, "Modeling, segmentation, and caliber estimation of bronchi in high-resolution computerized tomography," in *SPIE Statistical and Stochastic Methods in Image Processing II*, vol. 3167, 1997, Conference Proceedings, pp. 58–69.

[55] C. I. Fetita and F. J. Preteux, "Bronchial tree modeling and 3d reconstruction," in *SPIE Conference on Mathematical Modeling, Estimation and Imaging*, vol. 4121, 2000, Conference Proceedings, pp. 16–29.

[56] D. Aykac, G. Hoffman Ea Fau McLennan, J. M. McLennan G Fau Reinhardt, and J. M. Reinhardt, "Segmentation and analysis of the human airway tree from three-dimensional x-ray ct images," *IEEE Trans Med Imaging*, vol. 22, no. 0278-0062 (Print), pp. 940–50, 2003.

[57] A. P. Kiraly, G. Higgins We Fau McLennan, E. A. McLennan G Fau Hoffman, J. M. Hoffman Ea Fau Reinhardt, and J. M. Reinhardt, "Three-dimensional human airway segmentation methods for clinical virtual bronchoscopy," *Academic Radiology*, vol. 9, no. 1076-6332 (Print), pp. 1153–68, 2002.

[58] C. I. Fetita, F. Preteux, C. Beigelman-Aubry, and P. Grenier, "Pulmonary airways: 3-d reconstruction from multislice ct and clinical investigation," *Medical Imaging, IEEE Transactions on*, vol. 23, no. 11, pp. 1353–1364, 2004.

[59] A. Fabijanska, "Two-pass region growing algorithm for segmenting airway tree from mdct chest scans," *Computerized Medical Imaging and Graphics*, vol. 33, no. 7, pp. 537–546, 2009.

[60] Y. Sato, N. Nakajima S Fau Shiraga, H. Shiraga N Fau Atsumi, S. Atsumi H Fau Yoshida, T. Yoshida S Fau Koller, G. Koller T Fau Gerig, R. Gerig G Fau Kikinis, and R. Kikinis, "Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," *Medical Image Analysis*, vol. 2, no. 1361-8415 (Print), 1998.

[61] K. Krissian, G. Malandain, N. Ayache, R. Vaillant, and Y. Trousset, "Model-based detection of tubular structures in 3d images," *Computer Vision and Image Understanding*, vol. 80, pp. 130–171, 2000.

[62] Q. Li, S. Sone, and K. Doi, "Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional ct scans," *Medical Physics*, vol. 30, no. 8, pp. 2040–2051, 2003.

[63] C. Bauer and H. Bischof, *A Novel Approach for Detection of Tubular Objects and Its Application to Medical Image Analysis*, ser. Pattern Recognition Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2008, vol. 5096, book section 17, pp. 163–172.

[64] A. Frangi, W. Niessen, K. Vincken, and M. Viergever, *Multiscale vessel enhancement filtering*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 1998, vol. 1496, book section 14, pp. 130–137.

[65] M. W. Graham, J. D. Gibbs, D. C. Cornish, and W. E. Higgins, "Robust 3-d airway tree segmentation for image-guided peripheral bronchoscopy," *Medical Imaging, IEEE Transactions on*, vol. 29, no. 4, pp. 982–997, 2010.

[66] P. Jiantao, C. Fuhrman, W. F. Good, F. C. Sciurba, and D. Gur, "A differential geometric approach to automated segmentation of human airway tree," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 2, pp. 266–278, 2011.

[67] M. P. Fried, H. Moharir Vm Fau Shinmoto, A. M. Shinmoto H Fau Alyassin, W. E. Alyassin Am Fau Lorensen, L. Lorensen We Fau Hsu, R. Hsu L Fau Kikinis, and R. Kikinis, "Virtual laryngoscopy," *The Annals of Otology, Rhinology and Laryngology*, vol. 108, no. 0003-4894 (Print), pp. 221–6, 1999.

[68] C. Dongqing, L. Bin, P. Roche, H. Wei, C. Roque, and L. Zhengrong, "Feasibility studies of virtual laryngoscopy by ct and mri-from data acquisition, image segmen-

tation, to interactive visualization," *Nuclear Science, IEEE Transactions on*, vol. 48, no. 1, pp. 51–57, 2001.

[69] J. Liu, D. Udupa Jk Fau Odhnera, J. M. Odhnera D Fau McDonough, R. McDonough Jm Fau Arens, and R. Arens, "System for upper airway segmentation and measurement with mr imaging and fuzzy connectedness," *Academic Radiology*, vol. 10, no. 1076-6332 (Print), pp. 13–24, 2003.

[70] H. Shi, W. Scarfe, and A. Farman, "Upper airway segmentation and dimensions estimation from cone-beam ct image datasets," *International Journal of Computer Assisted Radiology and Surgery*, vol. 1, no. 3, pp. 177–186, 2006.

[71] E. Bresch and S. Narayanan, "Region segmentation in the frequency domain applied to upper airway real-time magnetic resonance images," *IEEE Trans Med Imaging*, vol. 28, no. 3, pp. 323–38, 2009.

[72] Z. Raeesy, S. Rueda, J. K. Udupa, and J. Coleman, "Automatic segmentation of vocal tract mr images," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, 2013, Conference Proceedings, pp. 1328–1331.

[73] D. Perperidis, R. Mohiaddin, and D. Rueckert, *Construction of a 4D Statistical Atlas of the Cardiac Anatomy and Its Use in Classification*, ser. Lecture Notes in Computer Science.   Springer Berlin Heidelberg, 2005, vol. 3750, book section 50, pp. 402–410.

[74] Z. Honghai, A. Wahle, R. K. Johnson, T. D. Scholz, and M. Sonka, "4-d cardiac mr image analysis: Left and right ventricular morphology and function," *Medical Imaging, IEEE Transactions on*, vol. 29, no. 2, pp. 350–364, 2010.

[75] M. van de Giessen, M. Foumani, F. M. Vos, S. D. Strackee, M. Maas, L. J. van Vliet, C. A. Grimbergen, and G. J. Streekstra, "A 4d statistical model of wrist bone motion patterns," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 3, pp. 613–625, 2012.

[76] A. I. A. Rahman, S.-H. Salleh, and K. A. K. Ahmad, "Analysis of vocal fold vibrations from high-speed digital images based on dynamic time warping," *International Journal of Medical, Health, Pharmaceutical and Biomedical Engineering*, vol. 8, no. 6, 2014.

[77] H. B. Logemann Ja Fau Fisher, B. Fisher Hb Fau Boshes, E. R. Boshes B Fau Blonsky, and E. R. Blonsky, "Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients," no. 0022-4677 (Print), 1978.

[78] C. Stewart, A. Winfield L Fau Hunt, S. B. Hunt A Fau Bressman, S. Bressman Sb Fau Fahn, A. Fahn S Fau Blitzer, M. F. Blitzer A Fau Brin, and M. F. Brin, "Speech dysfunction in early parkinson's disease," no. 0885-3185 (Print), 1995.

[79] A. V. Yuceturk, M. Yilmaz H Fau Egrilmez, S. Egrilmez M Fau Karaca, and S. Karaca, "Voice analysis and videolaryngostroboscopy in patients with parkinson's disease," no. 0937-4477 (Print), 2002.

[80] L. Perju Dumbrava, K. Lau, D. Phyland, P. Finlay, R. Beare, P. Bardin, P. Stuckey, P. Kempster, and D. Thyagarajan, "Vocal cords are hypokinetic in parkinson's disease," Departments of Neuroscience, Respiratory Medicine, Radiology Monash Medical Centre, Clayton, Victoria, Australia, Tech. Rep., 2014.

[81] G. Castellano, L. Bonilha, L. M. Li, and F. Cendes, "Texture analysis of medical images," *Clinical Radiology*, vol. 59, no. 12, pp. 1061–1069, 2004.

[82] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.

[83] B. S. Manjunath and R. Chellappa, "Unsupervised texture segmentation using markov random field models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 13, no. 5, pp. 478–482, 1991.

[84] P. Andrey and P. Tarroux, "Unsupervised segmentation of markov random field modeled textured images using selectionist relaxation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 3, pp. 252–262, 1998.

[85] J. M. Keller, S. Chen, and R. M. Crownover, "Texture description and segmentation through fractal geometry," *Computer Vision, Graphics, and Image Processing*, vol. 45, no. 2, pp. 150–166, 1989.

[86] J. Mao and A. K. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models," *Pattern Recognition*, vol. 25, no. 2, pp. 173–188, 1992.

[87] R. Chellappa and S. Chatterjee, "Classification of textures using gaussian markov random fields," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 4, pp. 959–963, 1985.

[88] J. Sun, D. Gu, S. Zhang, and Y. Chen, "Hidden markov bayesian texture segmentation using complex wavelet transform," *Vision, Image and Signal Processing, IEE Proceedings -*, vol. 151, no. 3, pp. 215–223, 2004.

[89] U. A. Ahmad, K. Kidiyo, and R. Joseph, "Texture features based on fourier transform and gabor filters: an empirical comparison," in *Machine Vision, 2007. ICMV 2007. International Conference on*, 2007, Conference Proceedings, pp. 67–72.

[90] S. Arivazhagan and L. Ganesan, "Texture classification using wavelet transform," *Pattern Recognition Letters*, vol. 24, pp. 1513–1521, 2003.

[91] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. SMC-3, no. 6, pp. 610–621, 1973.

[92] J. S. Weszka, C. R. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. SMC-6, no. 4, pp. 269–285, 1976.

[93] E. M. Darling and R. D. Joseph, "Pattern recognition from satellite altitudes," *Systems Science and Cybernetics, IEEE Transactions on*, vol. 4, no. 1, pp. 38–47, 1968.

[94] A. H. Mir, M. Hanmandlu, and S. N. Tandon, "Texture analysis of ct images," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 14, no. 6, pp. 781–786, 1995.

[95] L. Tesar, A. Shimizu, D. Smutek, H. Kobatake, and S. Nawano, "Medical image analysis of 3d ct images based on extension of haralick texture features," *Computerized Medical Imaging and Graphics*, vol. 32, no. 6, pp. 513–520, 2008.

[96] R. N. Sutton and E. L. Hall, "Texture measures for automatic classification of pulmonary disease," *Computers, IEEE Transactions on*, vol. C-21, no. 7, pp. 667–676, 1972.

[97] Y. P. Chien and F. King-Sun, "Recognition of x-ray picture patterns," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. SMC-4, no. 2, pp. 145–156, 1974.

[98] E. L. Hall, R. P. Kruger, and A. F. Turner, "An optical-digital system for automatic processing of chest x-rays," *Optical Engineering*, vol. 13, no. 3, pp. 133 250–133 250–, 1974, 10.1117/12.7971702.

[99] A. F. Turner, R. P. Kruger, and W. B. Thompson, "Automated computer screening of chest radiographs for pneumoconiosis," *Investigative Radiology*, vol. 11, no. 4, 1976.

[100] A. J. Coleman, K. A. Tonge, and S. C. Rankin, "The power spectral density as a texture measure in computed tomographic scans of the liver," *The British Journal of Radiology*, vol. 55, no. 656, pp. 601–603, 1982.

[101] U. Raeth, D. Schlaps, B. Limberg, I. Zuna, A. Lorenz, G. Van Kaick, W. J. Lorenz, and B. Kommerell, "Diagnostic accuracy of computerized b-scan texture analysis and conventional ultrasonography in diffuse parenchymal and malignant liver disease," *Journal of Clinical Ultrasound*, vol. 13, no. 2, pp. 87–99, 1985.

[102] R. F. Wagner, M. F. Insana, and D. G. Brown, "Unified approach to the detection and classification of speckle texture in diagnostic ultrasound," *Optical Engineering*, vol. 25, no. 6, pp. 256 738–256 738–, 1986, 10.1117/12.7973899.

[103] B. S. Garra, M. F. Insana, T. H. Shawker, R. F. Wagner, M. Bradford, and M. Russell, "Quantitative ultrasonic detection and classification of diffuse liver disease: Comparison with human observer performance," *Investigative Radiology*, vol. 24, no. 3, 1989.

[104] W. Chung-Ming, C. Yung-Chang, and H. Kai-Sheng, "Texture features for classification of ultrasonic liver images," *Medical Imaging, IEEE Transactions on*, vol. 11, no. 2, pp. 141–152, 1992.

[105] R. A. Lerski, K. Straughan, L. R. Schad, D. Boyce, S. Bluml, and I. Zuna, "Viii. mr image texture analysis - an approach to tissue characterization," *Magnetic Resonance Imaging*, vol. 11, no. 6, pp. 873–887, 1993.

[106] M. Ito, K. Ohki M Fau Hayashi, M. Hayashi K Fau Yamada, M. Yamada M Fau Uetani, T. Uetani M Fau Nakamura, and T. Nakamura, "Trabecular texture analysis of ct images in the relationship with spinal fracture," *Radiology*, no. 0033-8419 (Print), 1995.

[107] J. E. Koss, F. D. Newman, T. K. Johnson, and D. L. Kirch, "Abdominal organ segmentation using texture transforms and a hopfield neural network," *Medical Imaging, IEEE Transactions on*, vol. 18, no. 7, pp. 640–648, 1999.

[108] F. Chabat, G.-Z. Yang, and D. M. Hansell, "Obstructive lung diseases: Texture classification for differentiation at ct," *Radiology*, vol. 228, no. 3, pp. 871–877, 2003.

[109] D. S. Raicu, J. D. Furst, D. Channin, D.-H. U. I. Xu, A. Kurani, and S. Aioanei, "A texture dictionary for human organs tissues' classification," in *Proceedings of the 8th World Multiconference on Systemics, Cybernetics and Informatics*, 2004, Book.

[110] D.-H. Xu, A. S. Kurani, J. D. Furst, and D. S. Raicu, "Run-length encoding for volumetric texture," in *The 4th IASTED International Conference on Visualization, Imaging, and Image Processing*, 2004, Book.

[111] B. Podda and A. Giachetti, *Texture Analysis of CT Images for Vascular Segmentation: A Revised Run Length Approach*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2005, vol. 3617, book section 111, pp. 907–914.

[112] R. Khelifi, M. Adel, and S. Bourennane, "Spatial and spectral dependance co-occurrence method for multi-spectral image texture classification," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, 2010, Conference Proceedings, pp. 4361–4364.

[113] W. Gomez, W. C. A. Pereira, and A. F. C. Infantosi, "Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 10, pp. 1889–1899, 2012.

[114] T. Torheim, E. Malinen, K. Kvaal, H. Lyng, U. G. Indahl, E. K. F. Andersen, and C. M. Futsaether, "Classification of dynamic contrast enhanced mr images of cervical cancers using texture analysis and support vector machines," *Medical Imaging, IEEE Transactions on*, vol. 33, no. 8, pp. 1648–1656, 2014.

[115] K. Kwang In, J. Keechul, P. Se Hyun, and K. Hang Joon, "Support vector machines for texture classification," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 11, pp. 1542–1550, 2002.

[116] V. N. Vapnik, "An overview of statistical learning theory," *Neural Networks, IEEE Transactions on*, vol. 10, no. 5, pp. 988–999, 1999.

[117] M. A. Aizerman, E. A. Braverman, and L. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning," ser. Automation and Remote Control, vol. 6, 1964, Conference Proceedings, pp. 821–837, automation and Remote Control.

[118] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[119] L. C. Chang Chih, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2011.

[120] N. Otsu, "A threshold selection method from gray-level histograms," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 9, no. 1, pp. 62–66, 1979.

[121] S. Hu, E. A. Hoffman, and J. M. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric x-ray ct images," *Medical Imaging, IEEE Transactions on*, vol. 20, no. 6, pp. 490–498, 2001.

[122] M. T. Wu, A. A. Chang Jm Fau Chiang, J. Y. Chiang Aa Fau Lu, H. K. Lu Jy Fau Hsu, W. H. Hsu Hk Fau Hsu, C. F. Hsu Wh Fau Yang, and C. F. Yang, "Use of quantitative ct to predict postoperative lung function in patients with lung cancer," *Radiology*, no. 0033-8419 (Print), 1994.

[123] M. Saha, S. Agarwal, I. Arun, R. Ahmed, S. Chatterjee, P. Mitra, and C. Chakraborty, "Histogram based thresholding for automated nucleus segmentation using breast imprint cytology," in *Advancements of Medical Electronics*, S. Gupta, S. Bag, K. Ganguly, I. Sarkar, and P. Biswas, Eds. New Delhi: Springer India, 2015, pp. 49–57.

[124] l. C. Sefika Sahan, Ali Degirmenci, "A study on clustering based image thresholding techniques with matlab gui," *International Journal of Engineering Science and Computing*, vol. 6, no. 8, pp. 2152 – 2160, 2016.

[125] S. Manikandan, K. Ramar, M. W. Iruthayarajan, and K. Srinivasagan, "Multilevel thresholding for segmentation of medical brain images using real coded genetic algorithm," *Measurement*, vol. 47, pp. 558 – 568, 2014.

[126] Z.-K. Huang and K.-W. Chau, "A new image thresholding method based on gaussian mixture model," *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 899 – 907, 2008, special Issue on Advanced Intelligent Computing Theory and Methodology in Applied Mathematics and Computation.

[127] L. Liu, Z. Jia, J. Yang, N. Kasabov, and F. IEEE, "A medical image enhancement method using adaptive thresholding in nsct domain combined unsharp masking," *International Journal of Imaging Systems and Technology*, vol. 25, no. 3, pp. 199–205, 2015.

[128] M. E. Madden, *Introduction to Sectional Anatomy*.   Lippincott Williams & Wilkins, 2001.

[129] B. A. Ardekani, J. Kershaw, M. Braun, and I. Kanuo, "Automatic detection of the mid-sagittal plane in 3-d brain images," *Medical Imaging, IEEE Transactions on*, vol. 16, no. 6, pp. 947–952, 1997.

[130] S. Prima, S. Ourselin, and N. Ayache, "Computation of the mid-sagittal plane in 3-d brain images," *Medical Imaging, IEEE Transactions on*, vol. 21, no. 2, pp. 122–138, 2002.

[131] R. Zhang, T. Sato, and H. Arisawa, "Symmetry recognition using mid-sagittal plane extraction and tilt correction in 3d head images," in *SICE Annual Conference (SICE), 2013 Proceedings of*, 2013, Conference Proceedings, pp. 761–766.

[132] G. C. S. Ruppert, L. Teverovskiy, Y. Chen-Ping, A. X. Falcao, and L. Yanxi, "A new symmetry-based method for mid-sagittal plane extraction in neuroimages," in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, 2011, Conference Proceedings, pp. 285–288.

[133] C. Yuan, L. Wee Kheng, and L. Thiam Chye, "Automatic identification of frankfurt plane and mid-sagittal plane of skull," in *Applications of Computer Vision (WACV), 2012 IEEE Workshop on*, 2012, Conference Proceedings, pp. 233–238.

[134] A. G. Schwing and Y. Zheng, "Reliable extraction of the mid-sagittal plane in 3d brain mri via hierarchical landmark detection," in *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*, 2014, Conference Proceedings, pp. 213–216.

[135] M. B. P.Rajashekar Reddy, V.Amarnadh, "Evaluation of stopping criterion in contour tracing algorithms," *International Journal of Computer Science and Information Technologies*, 2012.

[136] J. K. Lee, *Computed body tomography with MRI correlation*. Lippincott Williams & Wilkins, 2006, vol. 1.

Appendices

Appendix A

Medical Image Visualiser Software

NifTI-1 images were used in our experiments. As explained in Section 1.4 these images consist of number of data volumes, and each volume consists of multiple axial slice images. Various different softwares like 3D slicer and ImageJ are available freely to view NifTI images. However, for the purpose of this research we developed a viewing software for NifTI images. Later, few additional functionalities were added to the viewer in order to use it for analysis during the research. In this appendix the details of this medical image visualising software will be explained.

Main window of the visualiser is shown in Figure A.1. Main window has four axes for the axial view, sagittal view, coronal view and the 3D view. Furthermore, on the right side of the window contains different panels which are dedicated for different functionalities such as file open, image colour/intensity adjustment, histogram generation, segmentation and cropping.
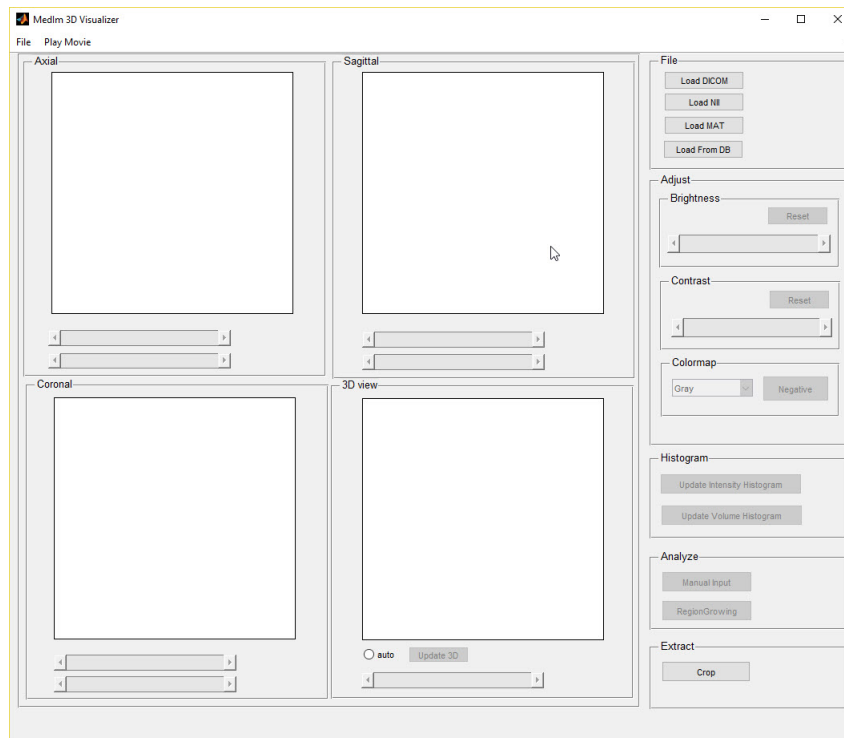
Figure A.1: Medical image visualiser main window

The visualiser software was originally intended to use by the radiologists or medical professionals who has a knowledge about the domain. However, some functionalities like viewing and colour adjustments might be useful to the normal users as well.

This software can handle DICOM, NifTI-1 or Matlab formatted data (*.mat*). To open a data volume the radiologists can use one of the corresponding Load button on File panel (Figure A.2(a)). Once the directory was given, the software will analyse and generate the different views (Figure A.2(b)) and display the images as shown in Figure A.3.
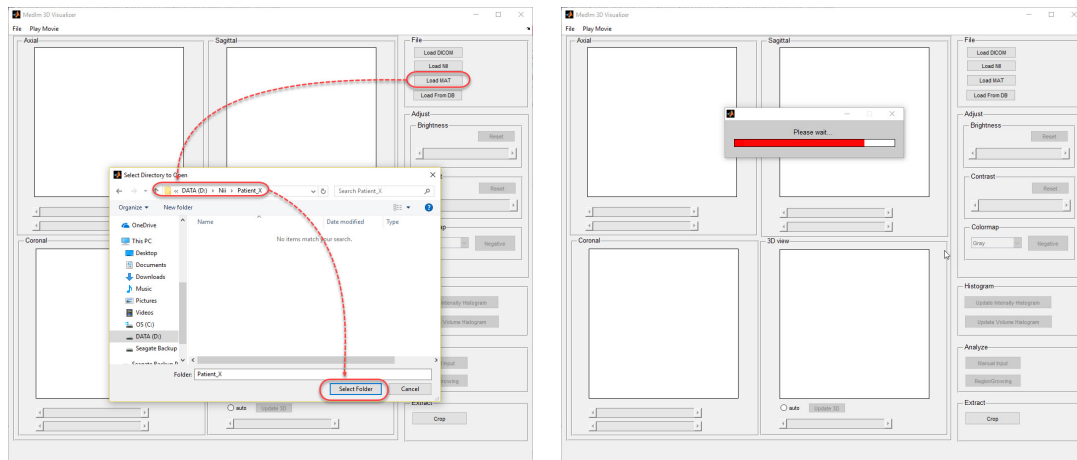
(a) Select 4D data directory *(b) Data is loading*

Figure A.2: Medical image visualiser: opening of a NifTY-1 data series
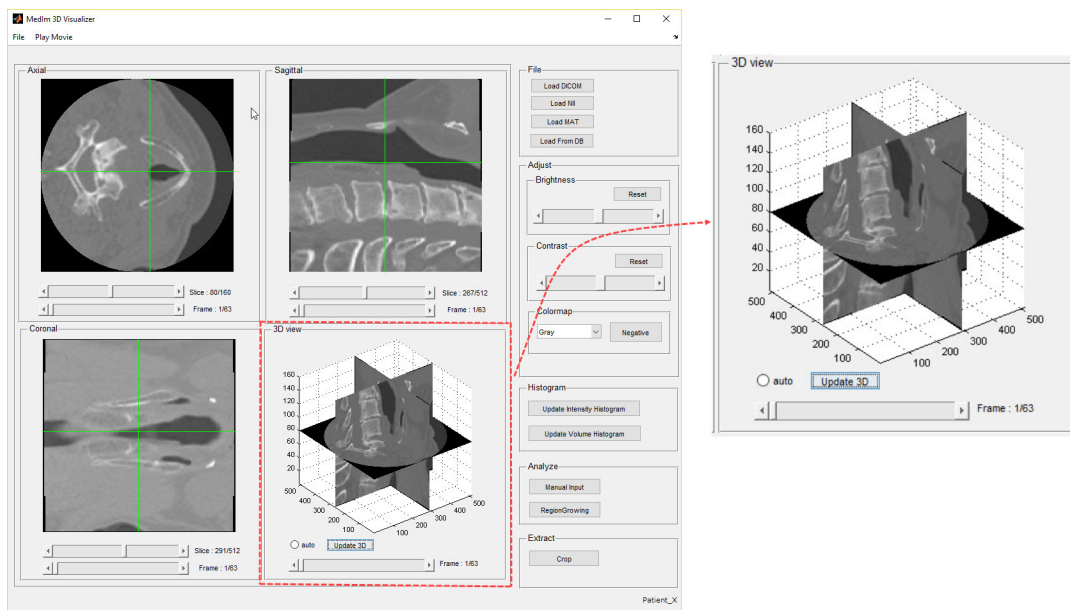
Figure A.3: Medical image visualiser: visualising of a NifTI-1 data series

Each Axial, Sagittal and the Coronal views has two sliders; one for changing the slice and the other for changing the frame. The 3D view is generated using the currently displayed three views (axial, sagittal and coronal) (Figure A.3). Users can check the auto radio button in the 3D view axes to allow automatic updating of the 3D view when any of the other views are changed. When the auto radio button is not checked, an update 3D button will be enabled, so that the users can update the 3D view whenever they require.

Adjust panel consists of three sliders to control brightness, contrast and colour maps of the images (Figure A.4(a)). The Histogram panel has two buttons to generate the histograms of axial, sagittal and coronal slices and the total data volume of the current frame (Figure A.4(b)).
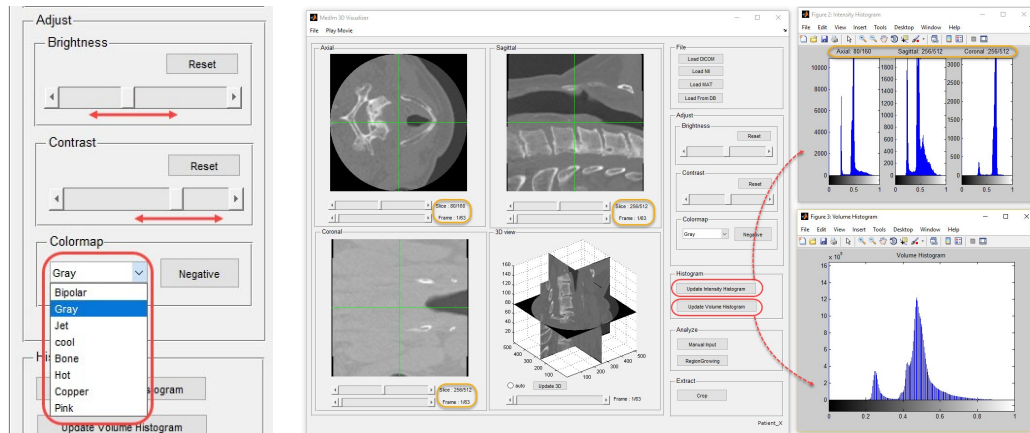
(a) Image colour / intensity adjustment panel

(b) Histograms generation

Figure A.4: Medical image visualiser: basic image processing interface

Our image visualiser contains the functionality to do manual segmentation and 2D region growing segmentation. Using the Manual Input button in Analyse panel the user can input a set of points in any of the axial, sagittal or coronal images to segment the required region (Figure A.5). Furthermore, we have implemented the region growing algorithm explained in Section 3.5.4 to perform the semi-automated segmentation of the airway on axial images. After pushing the RegionGrowing button on the Analyse panel, the user has to mark the initial seed point for the region growing segmentation on the axial image. Then the software will automatically segment the airway of that axial image, calculate the area and displays that on the GUI (Figure A.6).
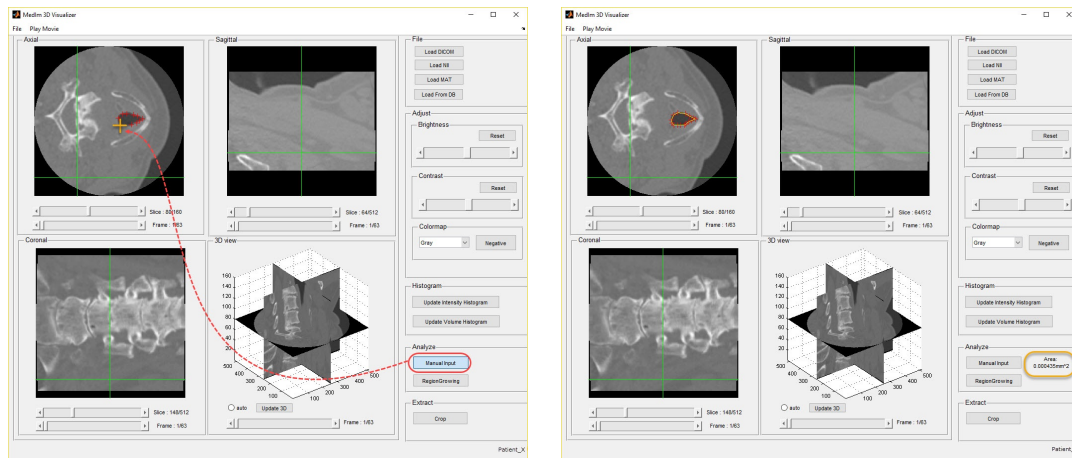
(a) Marking of the points   *(b) Manual segmentation result*

Figure A.5: Medical image visualiser: manual segmentation
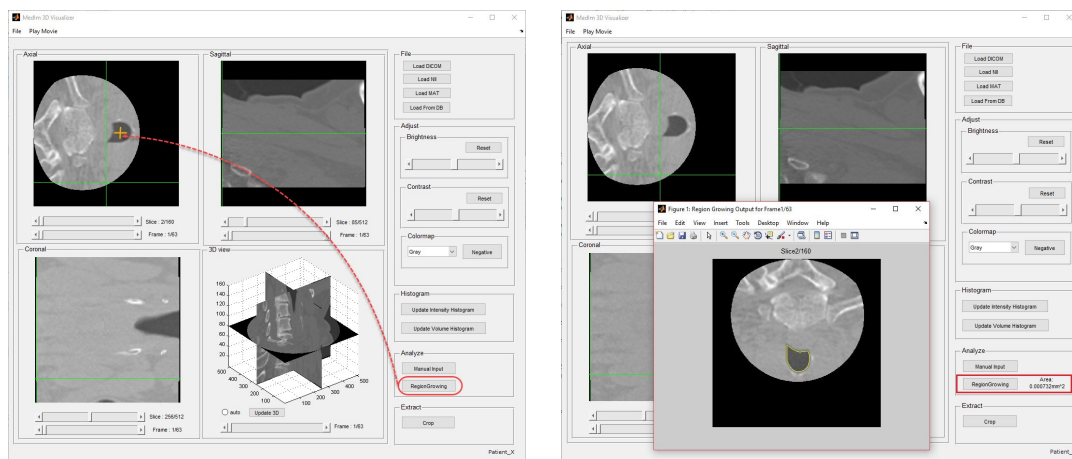
(a) Marking of seed point   *(b) Region growing result*

Figure A.6: Medical image visualiser: region growing segmentation

This image visualiser has the functionality to extract axial image patches from images as required. Crop button on Extract panel will display a movable/resizeable rectangle on the axial image, which allows the user to change the width, height and position. When the user double clicks on the rectangle, that image patch will be extracted and saved to the working directory (Figure A.7).
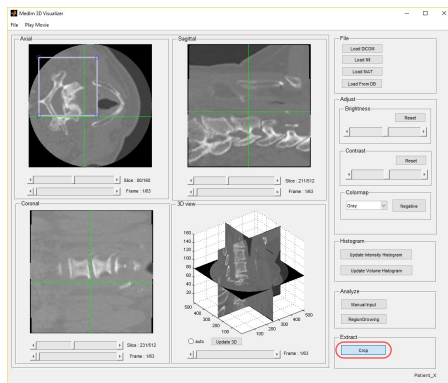
(a) Marking the position, height and width of the rectangle
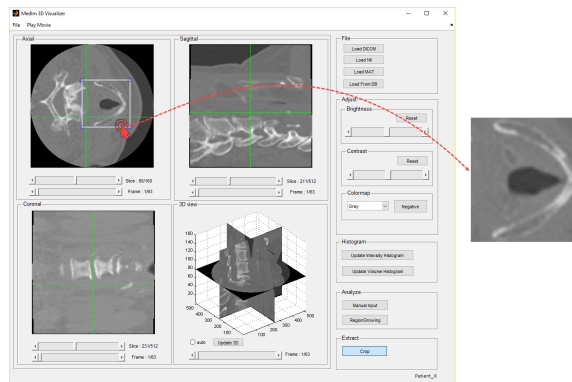
(b) Extracted patch

Figure A.7: Medical image visualiser: extracting image patches

Appendix B

Graphical User Interface for Vocal Folds Plane Estimation

In Chapter 4 the automated system developed as part of this research for estimating the vocal folds plane in 3D laryngeal CT images was discussed. At the end of that chapter we suggested the requirement of manual gold standard data to validate the output of the automated system. For this purpose, we developed a graphical user input software, so that the clinicians can easily validate the output of the vocal plane estimation system by confirming the correct decision or manually correcting the incorrect output. This appendix explains the details of this software and provides the instructions on how to use it.

This GUI was developed using Matlab 2014b Computer vision toolbox. The GUI is intended to be used by radiologists or clinicians who have knowledge about the problem. They can load laryngeal CT volumes which are in NifTI-1 file format to the software and the software will automatically run the vocal plane estimation algorithm discussed in Chapter 4 and display the estimated vocal plane output. Then the GUI will request the user to evaluate the accuracy of the decision, where the user can confirm the estimation or manually correct the location.

This software decreases the workload for clinicians by automatically estimating the vocal folds plane while giving the flexibility of manual intervention when required at the same time. Due to the necessity of high accuracies in decision-making with biomedical

data, the possibility of manual adjustment in automated systems is an essential requirement.

The main GUI window of our vocal folds plane estimation software is shown in Figure B.1(a). The radiologist can use the Open NII button on the main window to select the directory with NifTI-1 files of a particular patient (Figure B.1(b)). Then the software will analyse the NII files volume and computed MSP will be displayed (Figure B.2(a)). The estimated MSP is marked with a red line. Below the image there will be an instruction panel which asks the radiologist to confirm the choice of MSP of the data volume.

(a) Main window

(b) Select NII folder to open
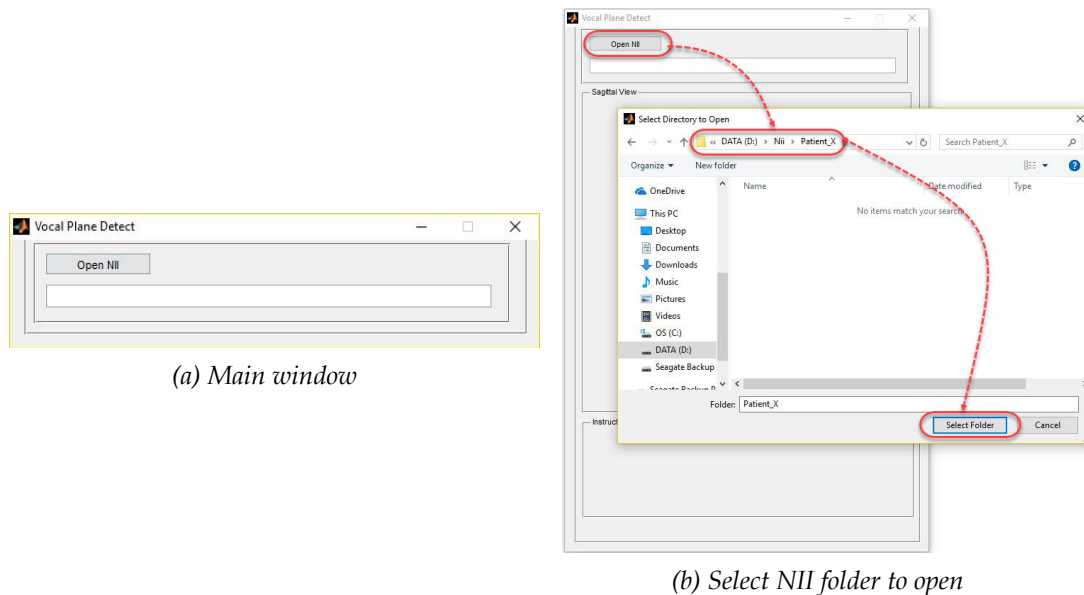
Figure B.1: Vocal plane estimation graphical user interface: opening of a new NII volume

The user can check the correctness of the MSP and select Yes or No accordingly. Selecting No will direct the user to the MSP adjusting window (Figure B.2(b)). The user can move the slider to choose the correct MSP and confirm.

(a) Mid-sagittal plane display   (b) Adjust the mid-sagittal plane
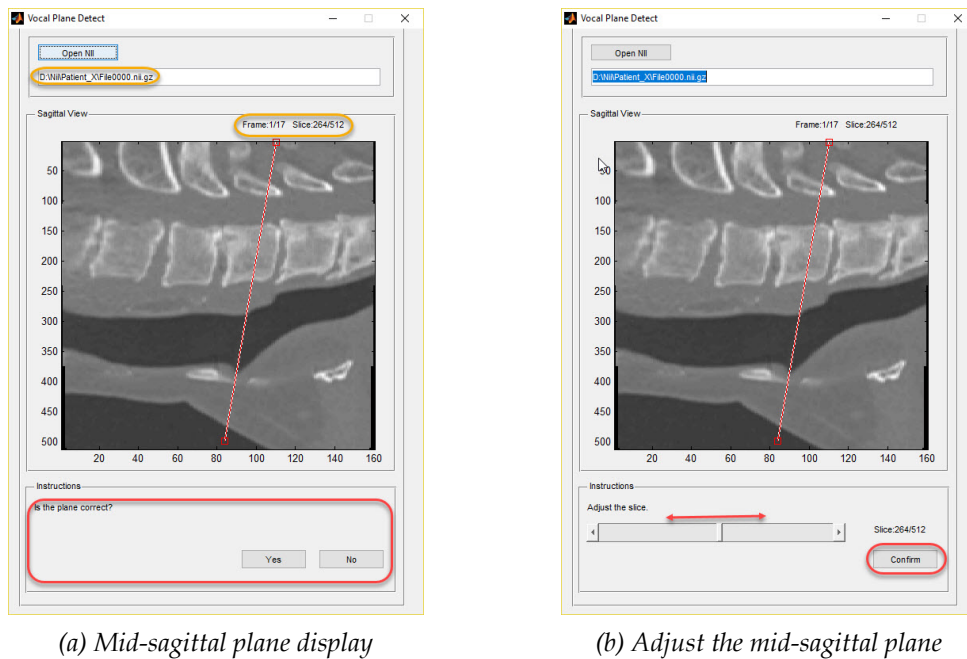
Figure B.2: Vocal plane estimation graphical user interface: mid-sagittal plane processing

Selecting Yes in MSP window (Figure B.2(a)) or Confirm in MSP adjusting window (Figure B.2(b)), both will direct the user to the vocal folds plane adjusting window (Figure B.3(a)). Here user can move the red line to reflect the correct vocal plane position and confirm. Finally, the adjusted position of the vocal folds plane is saved to a file and the MSP of the next frame of the data volume will be displayed for analysis (Figure B.3(b)).

All the manually adjusted vocal plane positions are recorded in the vocal_plane_locations.xls files, which can be found under the same directory as the GUI software. A sample record file is shown in Figure B.4.
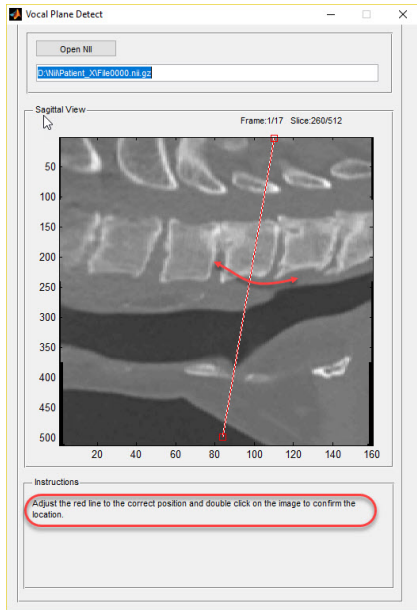
(a) Adjust the vocal plane position   *(b) Display next frame of the NII volume*
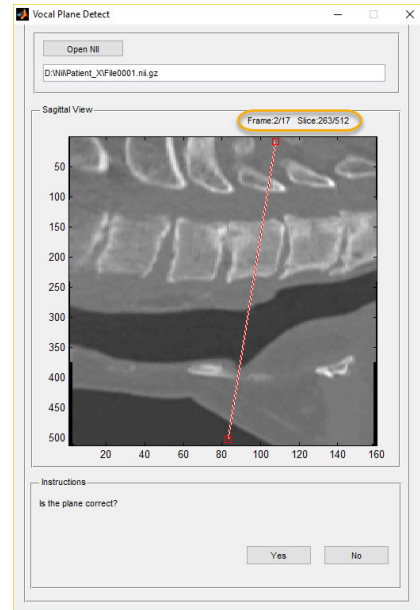
Figure B.3: Vocal plane estimation graphical user interface: vocal plane adjusting interface

Figure B.4: Vocal plane locations record

University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Author/s:
Hewavitharanage, Sajini Ruwanthika Gintota

Title:
Automatic analysis of 4D laryngeal CT scans to assist diagnosing of voice disorders

Date:
2017

Persistent Link:
http://hdl.handle.net/11343/198157

File Description:
Automatic analysis of 4D laryngeal CT scans to assist diagnosing of voice disorders