Universidade do Porto

FEUP Faculdade de Engenharia

Computational Processing and Analysis of Ear Images

Elisa Maria Lamego Barroso

July, 2011

Computational Processing and Analysis of Ear Images

Dissertation submitted to obtain the Master's degree in Biomedical Engineering

Elisa Maria Lamego Barroso

BSc in Biomedical Engineering by the School of Industrial Studies and Management, Polytechnic Institute of Porto (2009)

Supervisor:

Professor João Manuel R. S. Tavares

Assistant Professor of the Mechanical Engineering Department, Faculty of Engineering, University of Porto

Co-Supervisor:

Maria Fernanda Gentil Costa

Assistant Professor of the School of Health Technology of Porto, Polytechnic Institute of Porto

Acknowledgments

I would like to acknowledge my enormous debt to my advisor, Dr. João Manuel Tavares. His patience accurate insights helped me to become more technical in my academic research and make this practical work possible. He is more than an advisor, he is also a mentor. I would also like to thank to my co-supervisor Dr. Maria Fernanda Gentil.

I thank my father Manuel Barroso, mother Lurdes Barroso, and brother Alberto João, who have supported and loved me thought my life. I also need to thank my boyfriend Iúri Pereira, for his patience, understanding, and indomitable spirit, which provides an unfailing source of inspiration.

Finally, thank you to Yvonne Delayne for all the helps and supports. This would not have been possible without them.

Summary

The aims of this project was to study segmentation methods applied to medical images that allow the building of 3D geometric models of the inner ear and select the most efficiency for the segmentation of Computerized Tomography (CT) images.

The human auditory system belongs to a special senses group, which is characterized by having structures with highly localized receptors that provide specific information about the surrounding environment. This system consists of organs responsible for hearing and balance.

The human ear is divided into the outer ear, middle ear and inner ear. The latter consists of three main structures: the semicircular canals, the vestibule, which contains the utricle and saccule, and the cochlea.

In biomechanical studies of the inner ear are often used medical images obtained through different imaging techniques; in particular, Computerized Tomography (CT-standard, Micro-CT, Spiral-CT), Magnetic Resonance (MR-standard and Micro-MR) and even Microscopy of Histological Sections. Such images are used with the purpose of building 3D geometric models with realistic morphological characteristics and dimensions. For this goal, computational techniques were used in this project to analyze medical imaging; particularly, techniques of image processing and segmentation.

As the visualization of the inner ear, in the common medical images, is highly complex, the segmentation of the structures involved is often done manually; however, methodologies that allow the automatic segmentation of such images have been developed. In this project, these methodologies were analyzed taking into account the results obtained and the advantages and disadvantages of them were identified. Thus, it was identified the most efficient method to be applied to the segmentation of the inner ear structures.

Index

List of Figures

List of Tables

Acronyms

2D – Bidimensional

3D – Tridimensional

AAM – Active Appearance Model

AB – Appearance Based

ANN – Artificial Neuronal Network

ASM – Active Shape Model

CT – Computerized Tomography

D – Diagonal

DICOM – Digital Imaging and Communications in Medicine

EM – Estimation of the Maximum

FCM – Fuzzy C-means Algorithm

G – Gaussian

GAC – Geodesic Active Contour

GEM – Generalized Estimation of the Maximum

GGVFs – Generalized Version of the well-known Gradient Vector Flow snake

GVF – Gradient Vector Flow

H – Horizontal

ICP – Iterative Closest Point

ISODATA – Iterative Self-Organizing Data Analysis Technique Algorithm

KNN – K-nearest Neighbor

LS – Least Squares

MC – Marching Cube

ML – Maximum Likelihood

MR – Magnetic Resonance

MRC – Medical Research Council

PCA – Principal Components Analysis

PDM – Point Distribution Model

PVE – Partial-Volume Effect

RBFN – Radial Basis Function Neuronal Network

ROI – Region of Interest

SMC – Simplified Marching Cube

SVM – Support Vector Machine

US – United States

V – Vertical

WDM – Warp Distribution Model

Chapter I - Introduction to the Theme and Report Organization

- Introduction;
- Report Organization;
- Contributions.

1.1 – Introduction

The ear is a small but complex set of interlinked structures that are involved in both maintenance of normal balance and the sense of hearing. In order to hear, the ear collects the sound waves that arrive as air pressure variations and converts the waves into neurochemical impulses that travel along the cochlear-vestibular nerve to the brain (Seeley, Stephens et al. 2004; Irwin 2006). The organs of hearing and balance constitute the human auditory system, which can be divided into three main parts: external ear, middle ear and inner ear. The ear is by far the most complex organ of the human sensory system (Seeley, Stephens et al. 2004; Moller 2006).

The deafness is a hearing-impaired that is a severely disability and is considered a growing problem. The hearing loss can occur in one or both ears, and may be classified as mild, moderate, severe, or profound (Kaneshiro 2010). Nobody knows the exact number of hearing impaired people. However, Adrian Davis, of the British MRC Institute of Hearing Research estimates that the total number of people suffering from hearing loss superior to 25 dB will exceed 700 million by 2015. In 1995, there were 440 million hearing-impaired people in world-wide. In Europe, there were more than 70 million hearing-impaired people in a population of 700 million. The number of hearing-impaired people in North American is more than 25 million in a total population of 300 million (Davis 2010).

In every 1000 live births, about 2-3 infants will have some degree of hearing loss at birth. However, hearing loss can also develop in children who had normal hearing as infants (Kaneshiro 2010), because unilateral hearing loss is estimated to have a prevalence ranging between 0.1 and 5% in school-aged children. According to US government statistics (National Center for Health Statistics), between 1988 and 2006, mild or worse unilateral hearing loss (superior or equal to 25 dB), is informed in 1.8% of adolescents. Additionally, it was estimated 0.8% of mild or worse bilateral loss. In children, the hearing loss may maximize the loss of development of essential skills in speech, language, and social interactions (Melhem, Shakir et al. 1998; Hain 2010).

On the other hand, in older people, there is an epidemic hearing loss. The population aged 60 and older is hearing impaired, between 25 and 40%. Furthermore, hearing loss is also increasing with time and a hearing-impaired people have more trouble getting jobs, are paid less, and cannot communicate or enjoy music to the same extent as the rest of the population (Hain 2010).

All these values expose the real necessity of further studies in this area. Besides, the diagnosis and treatment of diseases of the middle and inner ear are made more difficult by the small size and hidden position of the associated organs in the temporal bone (Seemann, Seemann et al. 1999). Therefore, computational models have been used to simulate the behavior of the middle and inner ear in order to better understand the

relationship between its structure and function. Such as understand could aid to improve the design and function of prosthetics and the definition of better methodologies and plans in surgical procedures (Tuck-Lee, Pinsky et al. 2008).

In biomechanical studies of the ear, medical images that are obtained through different imaging techniques, Computerized Tomography (CT-standard, Micro-CT, Spiral-CT) (Christensen, He et al. 2003; Xianfen, Siping et al. 2005; Poznyakovskiy, Zahnert et al. 2008), Magnetic Resonance (MR-standard, Micro-MR) (Lane, Witte et al. 2005; Liu, Gao et al. 2007; Shi, Wang et al. 2010) and Histological Microscopy (Lee, Chan et al. 2010), have been often used. In these biomechanical studies are considered the use of guinea pig, cadaver, cat and chinchilla, for to characterize the biomechanical properties of the ear structures (Liu, Gao et al. 2007; Sim and Puria 2008). However, to build suitable biomechanical models, it is necessary to segment the ear structures in the images, previously acquired. The segmentation is the identification and separation of one or more structures in images.

The geometrical models to be used in biomechanical studies should present morphologies and dimensions similar to the real structures to be simulated. Usually, to obtain these models, the image segmentation is executed manually (Jun, Song et al. 2005; Liu, Gao et al. 2007; Tuck-Lee, Pinsky et al. 2008). The manual segmentation requires that the medical technicians outline the structure contours slice-by-slice by using pointing devices, such as a mouse or a trackball. This process is very time-consuming, and the results suffer from intra- or inter- observer variability. To answer the manual segmentation disadvantages, modern mathematical and physical techniques have been incorporated into computational algorithms. These incorporations have greatly enhanced the accuracy of the segmentation results (Yoo, Wang et al. 2001; Xianfen, Siping et al. 2005; Noble, Warren et al. 2008; Poznyakovskiy, Zahnert et al. 2008; Bradshaw, Curthoys et al. 2010; Shi, Wang et al. 2010)(Melhem, Shakir et al. 1998; Rodt, Ratiu et al. 2002; Christensen, He et al. 2003; Decraemer, Dirckx et al. 2003; Sim and Puria 2008; Comunello, Wangenheim et al. 2009; Lee, Chan et al. 2010) The computational segmentation algorithms can be classified into three essential classes: Thresholding, Clustering and Deformable Models. Frequently, techniques from different classes are combined in order to optimize the segmentation process (Ma, 2010).

Solutions of image processing and analysis are essential to attain realistic geometric models for the anatomical structures of the ear. Particularly, the segmentation of the ear structures in images is crucial to build patient-customized biomechanical models to be successfully used in computational simulations. From this simulation, the understanding of the connections between the ear structures and their functions becomes easier as well as the optimization of prosthetic implants.

The study and optimization of cochlear implant systems can be an important application area of the realistic and accurate modeling of the ear. In fact, the position of the

implanted electrodes has been identified as one of the most important variables in speech recognition, and the geometric modeling of the ear can facilitate the optimization of the electrode positions, which can be an essential step towards efficient traumatic cochlear implant surgeries. Up to now, only manual insertion tools or insertion aids exist, providing the possibility to insert the electrode using a fixed insertion technique that is not adjustable to the patient (Hussong, Rau et al. 2009; Rau, Hussong et al. 2010). Thus, based on accurate computational simulations the planning of surgical procedures can be enhanced (Tuck-Lee, Pinsky et al. 2008).

The biomechanical modeling of the ear also presents a key role in diagnosis and treatment of middle and inner ear diseases, because these two processes are hampered by the small size of the structures and by their hidden locations in the temporal bone (Seemann, Seemann et al. 1999). In addition, through the computational modeling of the inner ear, anatomical abnormalities of the bony labyrinth can be easier identified. Therefore, it is possible to create templates that standardize the abnormal configurations (Melhem, Shakir et al. 1998).

Extracting the structure contours in medical images, for example, by finding the image edges, can help doctors in detecting more efficient anomalies in visual inspections. However, the segmentation of structures in medical images is normally performed manually, requiring, for example, that medical technicians sketch the desired contours using pointing devices, such as a mouse or a trackball, which is very time-consuming and prone to errors. To overcome the disadvantages of manual segmentation, modern mathematical and physical techniques have been incorporated into the development of computational segmentation algorithms. These incorporations have greatly enhanced the accuracy of the segmentation results (Ma, Tavares et al. 2010).

Segmenting structures from medical images and reconstructing a compact geometric representation of these structures is difficult due to the sheer size of the datasets and the complexity and variability of the anatomic shapes of interest. Furthermore, the shortcomings typical of sampled data, such as sampling artifacts, spatial aliasing, and noise, may cause the boundaries of structures to be indistinct and disconnected. The challenge is to extract boundary elements belonging to the same structure and integrate these elements into a coherent and consistent model of the structure. Traditional low-level image processing techniques, which consider only local information, can make incorrect assumptions during this integration process and generate infeasible object boundaries. As a result, these model-free techniques usually require considerable amounts of expert intervention. Furthermore, the subsequent analysis and interpretation of the segmented objects is hindered by the pixel- or voxel-level structure representations generated by most image processing operations (McInerney and Terzopoulos 1996).

Image segmentation is a basic problem in image processing field and the key to the procedure from processing to analyzing. Extracting object contour of medical images,

obtaining specific edge information can help doctors understand diseases more visually and play an important role, especially for human vision (Xiao-Juan and Dan 2010).

Segmentation is made during automated analysis by delineating structures of interest and discriminating them from background tissue, and this separation process is generally effortless and swift for the human visual system. Therefore, the separation process performed by the human visual system can become a considerable challenge in algorithm development (Bankman 2000).

For the analysis of the medical images, segmentation is important for feature extraction, image measurements, and image-based diagnosis. For example, it may be useful to classify image pixels into anatomical regions, such as bones, muscles, and blood vessels, while in others into pathological regions, such as cancer, tissue deformities, and multiple sclerosis lesions (Bankman 2000).

The main objective of this work was to review image algorithms that have been used to segment the structures of the human ear. Hence, the identified algorithms will be analyzed, and their advantages and disadvantages will be pointed out, and some of their results will be presented and discussed. At the end of this dissertation the best method will be selected.

1.2 – Report Organization

This dissertation is structured into different and separate six chapters that are presented and summarized into five remainder sections:

- **Chapter II** – Anatomy and Physiology of the Ear
  In this chapter, a description of the ear is provided. Hence, the chapter starts with the explanation of the ear anatomy, followed by the description of the external, middle and inner ear. Afterwards, the chapter focuses on the physiologic events that happen in the ear: hearing and balance. Finally, the chapter presents the pathologies more frequent in this sensory organ. The main objective of this chapter is to understand the important physiological role and the importance of hearing and balance in society. This chapter also made possible the study of the shape and size of the structures that constitute the auditory and balance system.
- **Chapter III** – Medical Imaging Segmentation Algorithms
  This chapter describes the segmentation algorithms usually used in medical images and it was divided into three different sections. The first is named Algorithms based on Thresholding, the second includes Algorithms based on

Clustering and lastly the Algorithms based on Deformable Models are introduced. Through this chapter the segmentation algorithms and the main characteristics of each are expressed and studied.

- **Chapter IV** – Algorithms to Segment the Human Ear
  The chapter reviews algorithms that have used to segment the ear structures. These algorithms take into account the image type to be analyzed, the characteristics of the shape of the structures, the texture characteristics and the intensities range of the region of interest. This chapter is divided into three sections: Segmentation of Tympanic Membrane, Segmentation of Middle Ear, and Segmentation of Inner Ear, and includes applications of the algorithms reviewed in the segmentation of different ear structures.
- **Chapter V** – Experimental Results
  This chapter is organized into four sections, of which the first section presents a study about the existing segmentation techniques in medical images of the inner ear, in the second section is described a pre-processing technique for reducing noise and artifacts in the medical images. In the third section, the application of segmentation algorithms in the medical images of the ear is described. Finally, in the last section, the experimental results are discussed and the best algorithm to segment these human structures is selected.
- **Chapter VI** – Final Conclusions and Future Work
  Finally, this chapter presents the final conclusions and perspectives of future work.

1.3 - Contributions

The main contributions of this project can be highlighted as:

- The anatomic and physiology study of auditory and balance systems to a better understanding of the structures to be analyzed and to enhance the physiological relevance of them.
- The segmentation algorithms study was significant to know the wide variety of algorithms that can be used in segmentation of medical images and to assess the characteristics of each one.
- Segmentations performed on the CT images of the ear were essential for comparison and selection of the best segmentation method to be applied on this type of images.
- The reached conclusions contributed to indicate the best segmentation method and the pre-processing method that must be applied on CT images of the ear,

thus with this dissertation was realized a variety of studies that allow to understand the computer processing and analysis of the ear images.

Chapter II - Hearing: Anatomy, Physiology, and Clinical Focus of Ear Disorders

- Introduction;
- Anatomy;
- Physiology;
- Clinical Focus of Ear Disorders;
- Summary

2.1 Introduction

The human auditory system is one of the special senses that are defined by including structures that have highly localized receivers, which provide specific information about the surrounding environment (Seeley, 2004).

As a sensory organ, the ear is the most vital, because the hearing loss may maximize the loss of development of essential skills in speech, language, and social interactions (Melhem, 1998).

In this chapter, a description of the ear is provided. Hence, the chapter starts with the explanation of the ear anatomy, followed by the description of the external, middle and inner ear. Afterwards, the chapter focuses the physiologic events that happen in the ear: hearing and balance. Finally, the chapter presents the pathologies more frequents in this sensory organ.

2.2 Anatomy

In humans, the ear percepts and interprets the sound waves in a frequency range from 16 Hz to 20 kHz (intensity range: 0 – 130 dB) (Henrique 2002).

The ear includes three main parts: the outer ear, the middle ear and the inner ear, Figure 2.1. The external ear includes the auricle and the external auditory meatus. The external ear terminates medially at the eardrum, or tympanic membrane. Moreover, the middle ear is an air-filled space within the petrous portion of the temporal bone. In the air-filled space there are the auditory ossicles. The vestibular apparatus and the cochlea belong to the inner ear, being the first structure responsible for the balance function and the second one for hearing (Moller 2006).

2.2.1 Outer Ear

The outer ear is characterized by three structures: the auricle, the external auditory meatus, and the tympanic membrane. In older individuals, the size of the outer ear increases, especially in men (Moller 2006).

The auricle consists primarily of elastic cartilage covered by skin, Figure 2.2, and the shape of the auricle helps to collect sound waves and direct them toward the external auditory meatus (Seeley, Stephens et al. 2004).

The different parts of the auricle have specific names. The concha, which is a part of the auricle, is acoustically the most important.

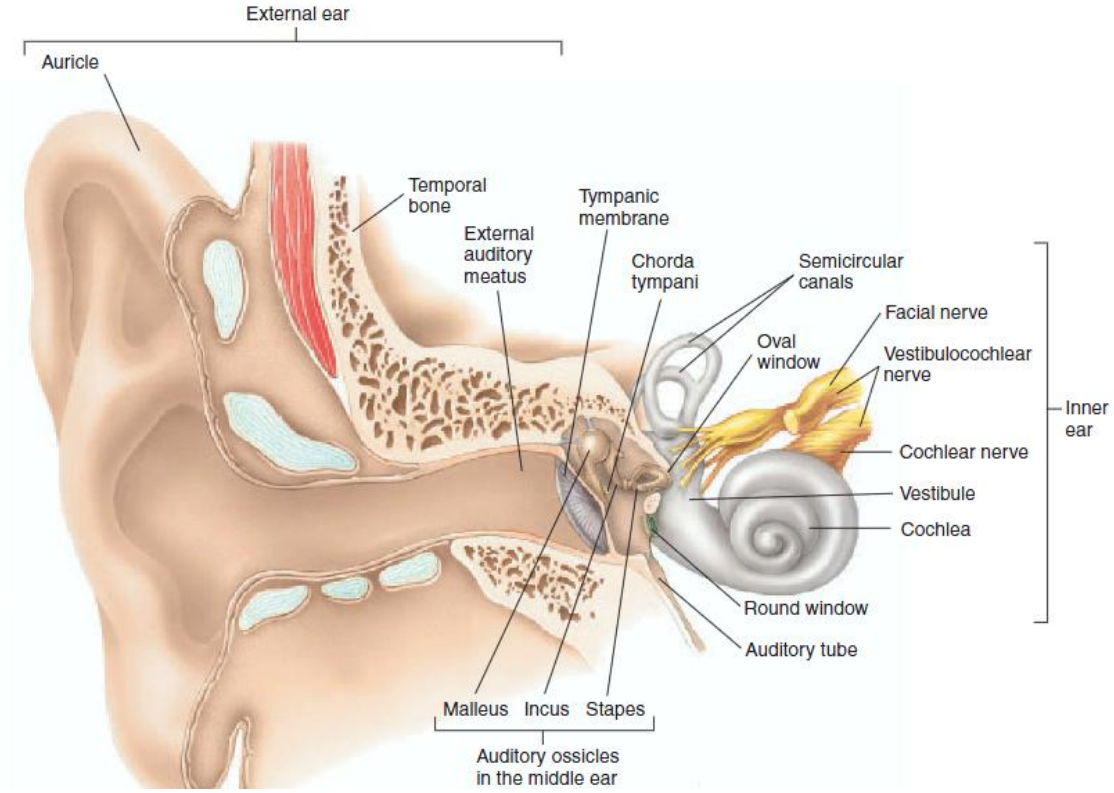


Figure 2.1: External, middle and inner ear (from (Seeley, Stephens et al. 2004)).

The external auditory meatus converges in the ear canal, which has a shape similar to a lazy "S", a length approximately 25 mm and a diameter around 6 mm. A nearly circular opening in the skull bone represents the most medial part. The outer part is cartilage, and the outer cartilaginous portion is also nearly circular in young individuals, but with age, the cartilaginous part often changes shape. In addition, the lumen of the ear canal frequently becomes smaller.

The ear canal is covered by skin that secrets cerumen and has hairs on its surface. There are two types of cells that contribute to secretion of cerumen: sebaceous cells, located close to the hair follicles, and the ceruminous glands. The sebaceous glands form their secretion by passive breakdown of cells because they cannot secrete actively. There are two kinds of cerumen: dry and wet. The overproduction of cerumen may block the meatus.

The tympanic membrane separates the external ear from the middle ear and is a thin, semitransparent, nearly oval, three layered membrane. It is cone-shaped, with a height of 2 mm with the apex pointed inward (Seeley, Stephens et al. 2004; Moller 2006).

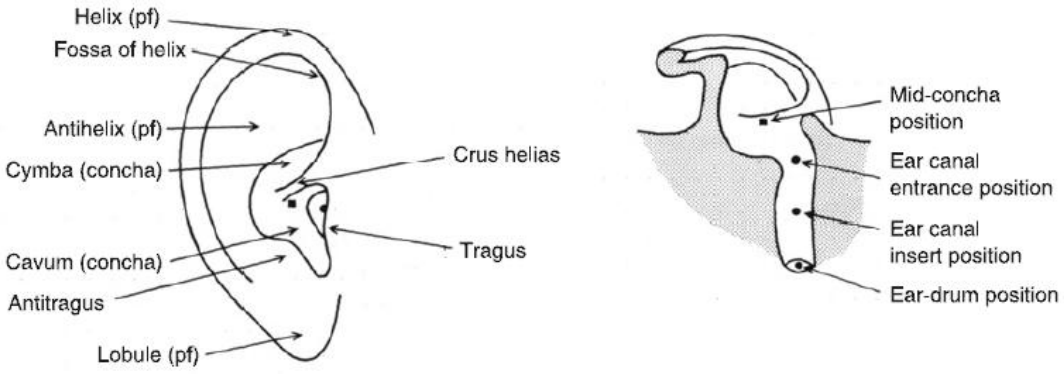


Figure 2.2: Human external ear (from (Moller 2006)).

2.2.2 Middle Ear

The inner face of the tympanic membrane is integrated in the middle ear. Furthermore, the middle ear has three bones (ossicles): the malleus, the incus and the stapes, Figure 2.3. The tensor tympani muscle and the stapedius muscle are two small muscles that are also located in the middle ear. The manubrium of malleus is imbedded in the tympanic membrane, and the head of malleus is connected to the incus that in turn connects to the stapes. The footplate is located in the oval window of the cochlea and is a part of stapes.

The facial nerve or the nervous intermedius travels across the middle ear cavity through the chorda tympani that is a branch of the facial nerve.

The tympanic membrane, can be seen from the ear canal. It is slightly concave and has a surface of about 85 mm^2, Figure 2.4. The main part of the tympanic membrane is the pars tensa, which has an area of approximately 55 mm^2, and composed of radial and circular fibers overlaying each other. These fibers provide a lightweight stiff membrane that is ideal for converting sound into the vibration of the malleus. Furthermore, the fibers have mechanical properties to recognize vibration because they are comprised of collagen. The pars flaccid are a smaller part of the tympanic membrane and are located above the manubrium of malleus. This smaller part is thicker than the pars tensa, but its fibers are not arranged as orderly as the collagen fibers of the pars tensa.

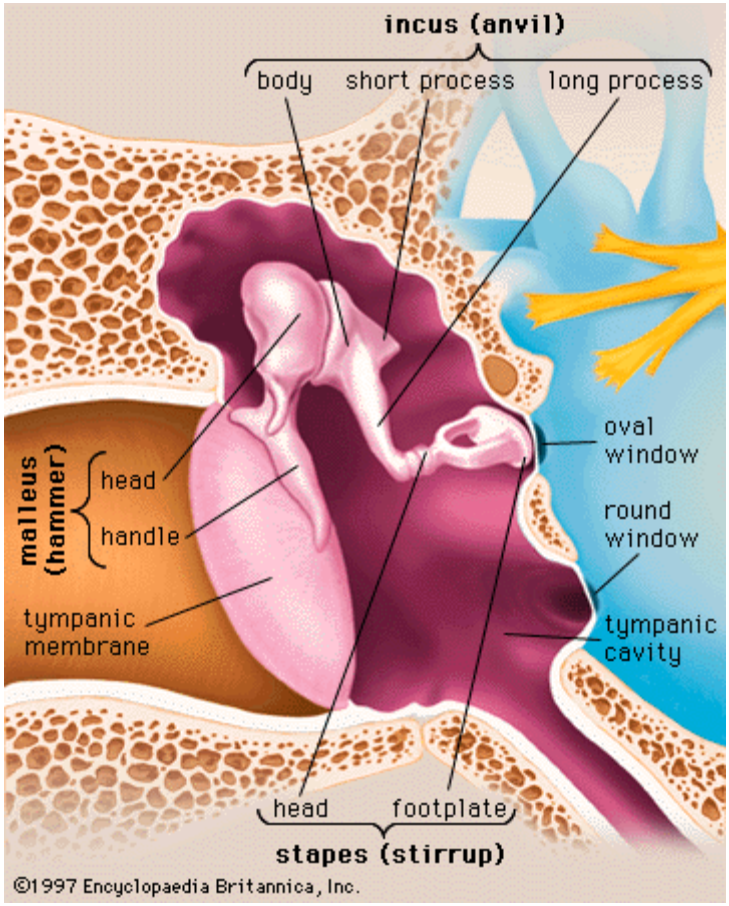


Figure 2.3: Middle ear (from (Carr 2010)).

The middle ear contains three auditory ossicles that are suspended by several ligaments: the malleus, incus, and stapes, Figure 2.5.

The tip of the manubrium of the malleus is located in the apex of the tympanic membrane. The head of the malleus is suspended in the epitympanum. The short process of the incus is held in place by the posterior incudal ligament and rests in the fossa incudo of the malleus. The long process of the incus forms one side of the incudo-stapedial joint. The head of the malleus and the incus are linked by a double saddle joint. The joint between these two bones appears rigid, but it allows flexibility for the movements of the stapes. The stapes is induced by contraction of the stapedius muscle. Furthermore, the stapes is suspended in the oval window of the cochlea buy two ligaments.

Two small muscles are located in the middle ear: the tensor tympani muscle and the stapedius muscle.

The tensor tympani muscle is attached to the manubrium of the malleus and extends between the malleus and the wall of the middle ear cavity (near the entrance to the Eustachian tube). The tensor tympani muscle is innervated by the trigeminal nerve.

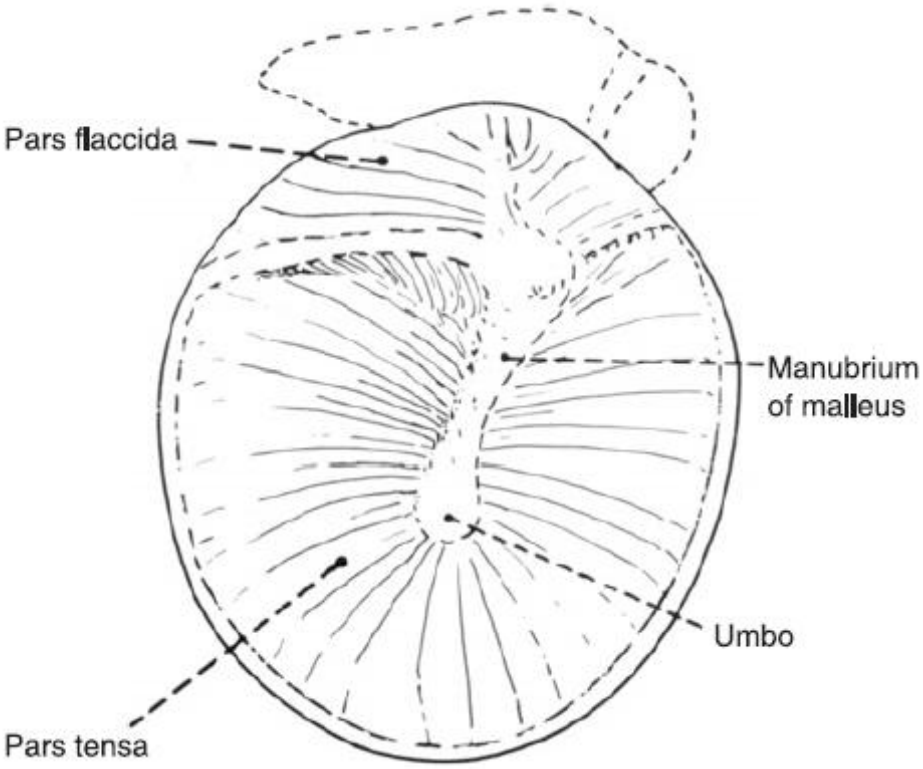


Figure 2.4: The tympanic membrane (from (Moller 2006)).

The stapedius muscle is attached to the head of the stapes, and most of the muscle is located in a bony canal. The stapedius muscle is considered the smallest striate muscle of the body. The facial nerve is responsible for the innervate of the stapedius muscle (Seeley, Stephens et al. 2004; Moller 2006).

The Eustachian tube is located in the inner ear cavity, Figure 2.6. The tube consists of a bony part (the protympani), and is responsible for keeping air pressure in the middle ear cavity close to the ambient pressure. The Eustachian tube is 3.5 to 3.9 cm long and presents approximately 45 degrees relatively to the horizontal direction. In young children, the Eustachian tube is shorter and is directed nearly horizontally. The Eustachian tube opens due to contraction of the tensor veli palatini muscle, Figure 2.7. This muscle is located in the pharynx and innervated by the motor portion of the fifth cranial nerve. Positive air pressure can open the Eustachian tube, but negative pressure may close it harder.

The tympanum, the epitympanum and the system of the mastoide air cells are the middle ear cavities. The tympanum is the main cavity that lies between the tympanic membrane and the wall of the inner ear, which is called promontorium.

The epitympanum is a smaller part located between the tympanum and the head of malleus. The middle ear cavity presents a volume around 2 cm^3, but this value varies considerably from person to person. If the volume of mastoide cells is considered, the total volume can be as large as 10 cm^3 (Seeley, Stephens et al. 2004; Moller 2006).

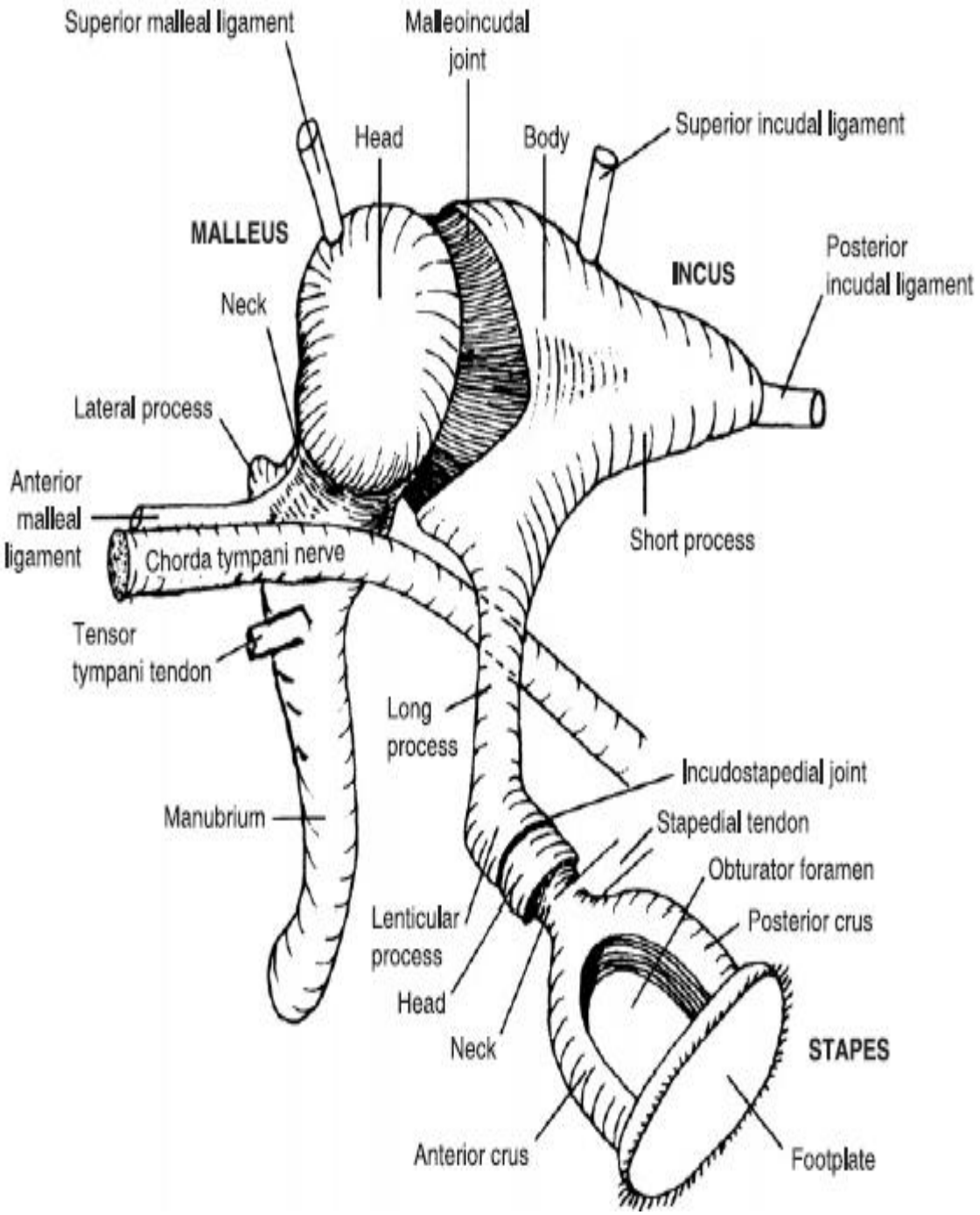


Figure 2.5: Muscles and ossicles of the middle ear (from (Moller 2006)).

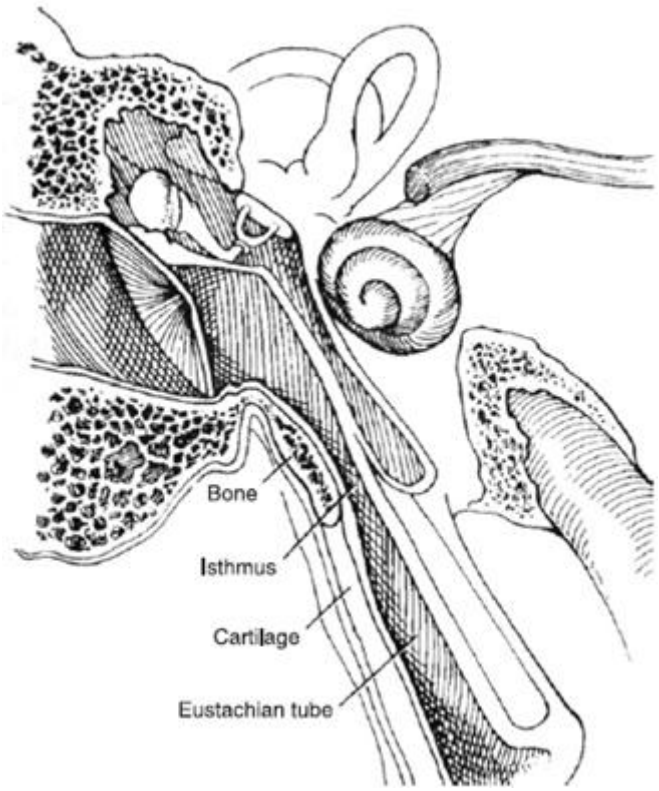


Figure 2.6: Middle ear cross-section showing the Eustachian tube (from (Moller 2006)).

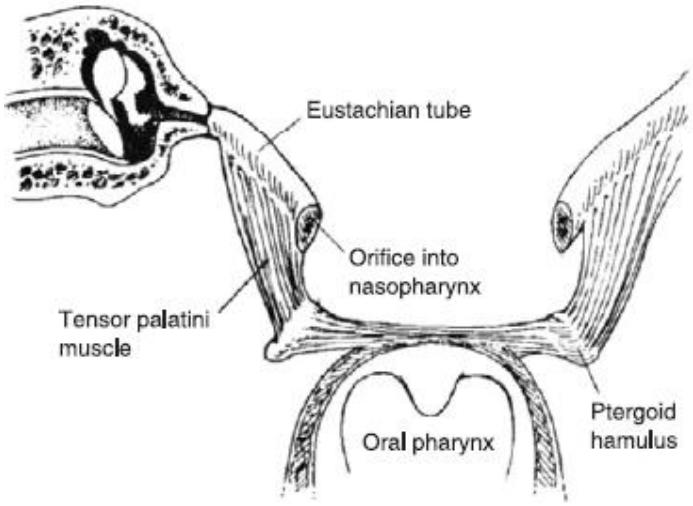


Figure 2.7: Orientation of the Eustachian tube and localization of the tensor veli platini muscle (from (Moller 2006)).

2.2.3 Inner Ear

The inner ear contains tunnels and chambers that are located inside the temporal bone, Figure 2.8, and known as the bony labyrinth. The bony labyrinth is lined with the periosteum and has a membranous labyrinth inserted.

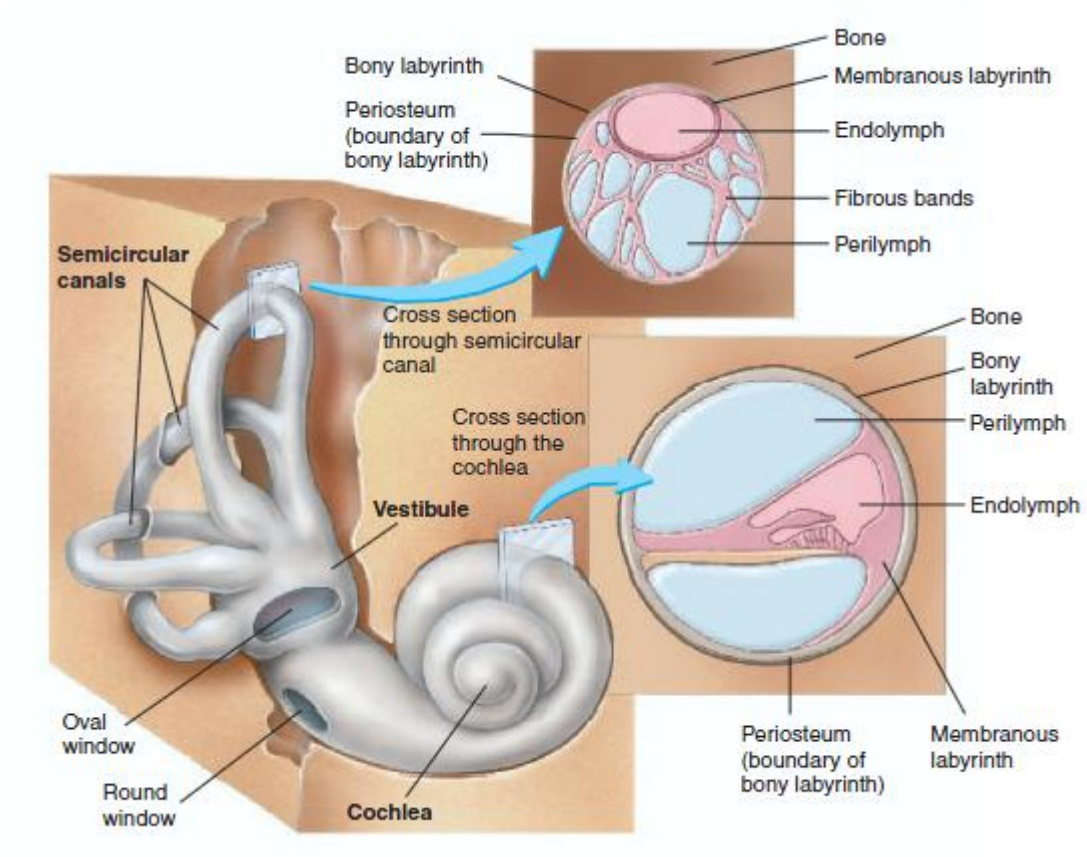


Figure 2.8: The bony and membranous labyrinths of the inner ear (from (Seeley, Stephens et al. 2004)).

The membranous labyrinth is filled with a clear fluid called endolymph. The space between the membranous and the bony labyrinth is also filled by a fluid known as perilymph. The perilymph is very similar to cerebrospinal fluid,but the endolymph has a high concentration of potassium and low concentration of sodium; therefore, the concentration of endolymph fluid is opposite to that of the concentration of the perilymph fluid.

The bony labyrinth is divided into three regions: cochlea, vestibule, and semicircular canals. The cochlea is the structure responsible for the hearing function. The vestibule and the semicircular canals are involved in the balance.

The membranous labyrinth of the cochlea is divided into three parts: the scala vestibuli, the scala tympani, and the scala media, Figure 2.9. The oval window communicates with the vestibule, which in turns connects with the scala vestibuli. The scala vestibuli extends from the oval window to the helicotrema at the apex of the cochlea whereas the scala tympani extends from the helicotrema (the area of aperture is approximately 0.05 mm^2 in humans), back from the apex and is parallel to the scala vestibuli, to the membrane of the round window. The perilymph is inserted in these two chambers, the scala vestibuli and tympani. The scala media is separated from the scala vestibuli by Reissner's membrane and from the scala tympani by the basilar membrane.

The fluid of scala media is an ionic composition similar to an intracellular fluid that is rich in potassium and low in sodium, but the fluid in the scala vestibuli and in the scala tympani is similar to the extracellular fluid. The basilar membrane separates sounds according to their frequency, and along the basilar membrane can be found the organ of Corti. There are many different kinds of cells in the organ of Corti. The hair cells, i.e. the sensory cells, are arranged in rows along the basilar membrane and can be of two main types: the outer and the inner hair cells, Figures 2.10A and B, respectively.

The human cochlea has approximately 12000 outer hair cells that are arranged in 3-5 rows along the basilar membrane. It is estimated that approximately 3500 inner hair cells, arranged in a single row exist. The inner hair cells are different from the outer hair cells, one of the reasons being different shapes. While the inner hair cells are flask-shaped, the outer hair cells are cylindrical. In the apical region of the cochlea the outer hair cells are longer than in the more basal regions, approximately 8 μm long in the apical region and less than 2 μm in the base.

The diameter of the longest outer hair cells is approximately one tenth of the human hair cells diameter. Inner hair cells have the same dimension in the entire cochlea. Hence, the hair cells have hairlike projections at their apical ends and these projections, in children, consist of one cilium (kinocilium) and about 80 very long microvili, often referred to as stereocilia. In adults, the cilium is absent from most hair cells. The stereocilia are located at the base of the cochlea in the inner hair cells. All inner hair cells have approximately the same number of stereocilia (approximately 60) and whereas each outer hair cell has 50-150 stereocilia and are arranged in 3-4 rows that assume a W or V shape.

Other types of cells that are found in the cochlea are the Deiter's cells and Henson's cells, which support cells of the organ of Corti.

One important structure that is located between the perilymphatic and the endolymphatic space along the cochlear wall is the stria vascularis. This structure has a

rich blood supply and its cells are rich in mitochondria. The presence of mitochondria indicates that it is involved in metabolic activity.

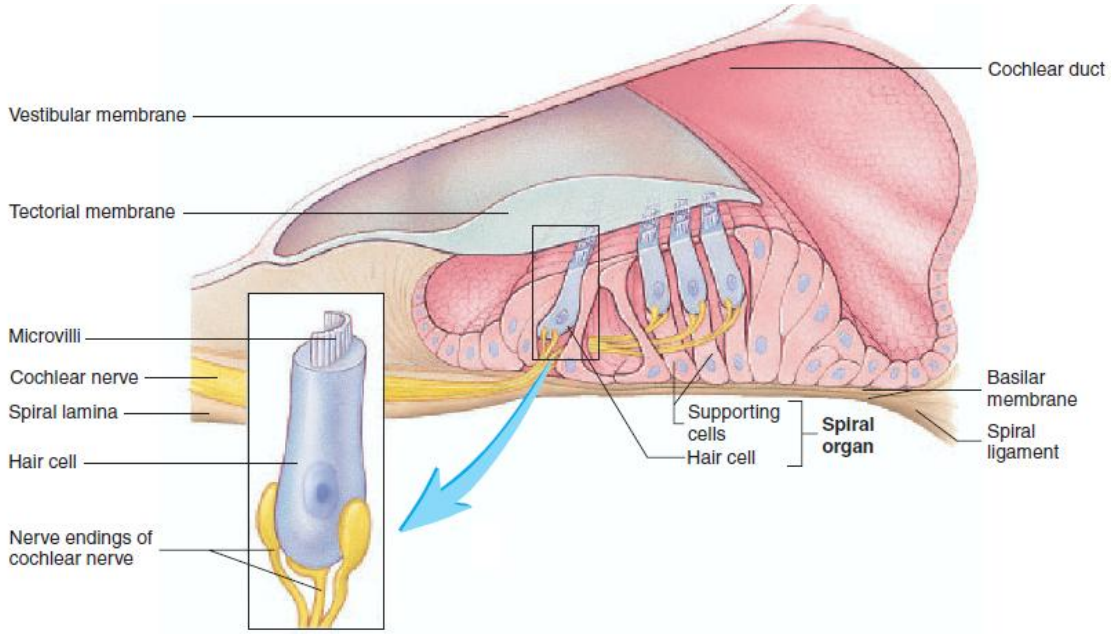


Figure 2.9: An enlarged section of the cochlear duct (membranous labyrinth) and a greatly enlarged individual sensory hair cell (from (Seeley, Stephens et al. 2004)).

The basilar membrane is a connective tissue forming the floor of the scala media. In the base of the cochlea, the basilar membrane has a width of approximately 150 µm. The basilar membrane is about 450 µm wide at the apex and is stiffer in the basal end than at the apex.

The cochlea is innervated by three types of nerve fibers: afferent auditory nerve fibers, efferent nerve fibers and automatic nerve fibers. The first type, afferent auditory nerve fibers are bipolar cells, the cell bodies of which are located in the spiral ganglion (located in a bony canal – the Rosenthal's canal). The auditory nerve has approximately 30000 afferent nerve fibers. There are two types of afferent fibers: Type I and Type II. Type I auditory nerve fibers are myelinated, have large cell bodies and comprise 95% of the auditory nerve fibers. On the other hand, Type II represents approximately 5% of the auditory nerve, is unmyelinated and has small cell bodies. Many Type I auditory nerve fibers terminate on each inner hair cell while a single Type II auditory nerve fiber connects with many outer hair cells. Each inner hair cell receives approximately 20 nerve fibers. The efferent auditory fibers, also called olivocochlear bundle, receive different connections from the descending auditory nervous system. Humans have approximately 500-600 efferent fibers that have their cell bodies in the nuclei of the superior olivary complex (SOC) of the brain stem. The efferent fibers are of two kinds: medial olivocochlear fibers and lateral olivocochlear efferent fibers. Medial olivocochlear fibers are large myelinated fibers that originate in the medial superior

olivary complex. This kind terminates on outer hair cells. Each outer hair cell receives many efferent fibers and each efferent fiber connects with several outer hair cells. The lateral olivocochlear fibers are small unmyelinated fibers that originate in the lateral nucleus of the superior olivary complex. Efferent fibers connect more sparsely to inner hair cells. The autonomic fibers are also responsible for the autonomic nerve supply in the inner ear. These fibers mainly innervate blood vessels; however, they also contact hair cells (Seeley, Stephens et al. 2004; Moller 2006).

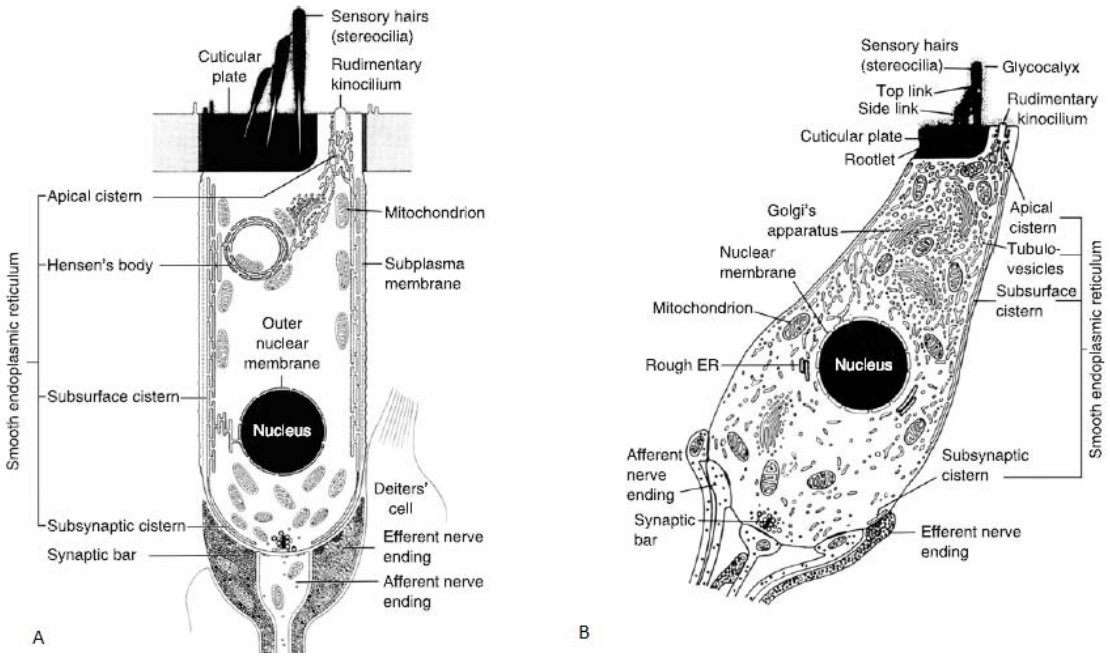


Figure 2.10: Schematic drawing of the cross-section of an outer hair cell (A) and of an inner hair cell (B) (from (Moller 2006)).

2.3 Physiology

The ear is involved in the maintenance of normal balance and sense of hearing. In order to hear, the ear collects the sound waves that arrive as pressure changes in air, and converts these into neurochemical impulses that travel along cochlear-vestibular nerve to the brain. There are both active and passive mechanisms involved in this process. The prime function of the vestibular system is to detect and compensate for movement. This includes the ability to maintain optic fixation despite movement and to initiate muscle reflexes to maintain balance.

For the purposes of describing the function of the ear it is usually divided into four different parts: outer ear, the middle ear and the auditory and the vestibular parts of the inner ear.

2.3.1 Auditory Function

Vibration of matter such as air, water or a solid material creates sound. No sound occurs in a vacuum. When a person speaks, the vocal cords vibrate, causing the air passing out of the lungs to vibrate. The vibrations consist of bands of compressed air followed by bands of less compressed air. These vibrations are propagated through the air as sound waves, somewhat like ripples are propagated over the surface of water (Moller 2006; Martin and Clark 2011).

Pitch is a function of the wave frequency measured in hertz. The higher the frequency of a waveform, the higher the pitch of the sound you hear. The normal range of the human hearing is 20-20000 Hz and 0 (zero) or more decibels. Therefore, sounds louder than 125 db are painful to the ear. The range 250-8000 Hz is the range that is tested for the possibility of hearing impairment because it is the most important for communication (Seeley, Stephens et al. 2004).

Timbre is the resonance quality or overtones of a sound. A smooth sigmoid curve is the image of a "pure" sound wave, but such a wave almost never exists in nature (Seeley, Stephens et al. 2004). The steps involved in hearing are the following:

- The auricle collects sound waves that are then conducted through the external auditory meatus to the tympanic membrane, causing it to vibrate;
- The vibrating tympanic membrane causes the malleus, incus and stapes to vibrate;
- Vibration of the stapes produces vibration in the perilymph of the scala vestibule;
- The vibration of the perilymph produces simultaneous vibration of the vestibular membrane and the endolymph in the cochlear duct;
- Vibration of the endolymph causes the basilar membrane to vibrate;
- As the basilar membrane vibrates, the hair cells attached to the membrane move relative to the tectorial membrane, which remains stationary;
- The hair cell microvilli, embedded in the tectorial membrane become bent;
- Bending of the microvilli causes depolarization of the hair cells;

- The hair cells induce action potentials in the cochlear neurons;
- The action potentials generated in the cochlear neurons are conducted to the CNS;
- The action potentials are translated in the cerebral cortex and are perceived as sound.

a) External Ear

The auricle collects sound waves that are conducted through the external auditory meatus toward the tympanic membrane. The various folds in the pinna's structure amplify some high-frequency components of the sound. They also help in the localization of sound in the vertical plane. As sounds hit the pinna from above and below, their paths to the external auditory meatus vary in length. This means that they take different times to reach the meatus. Again, this is a feature of high-frequency sounds. The difference in time of arrival of the low-frequency and high-frequency components of the sounds allows for localization. The pinna is also involved in localization of sound from in front and behind. As sound waves pass the pinna from behind, they are diffracted around the pinna to the meatus whereas sounds from in front do not do this. The slight distortion produced allows for localization (Seeley, Stephens et al. 2004; Moller 2006; Martin and Clark 2011).

Localization of sound in the lateral plane is a function of the pinnas being on different sides of the head. A sound directly from the left reaches the left ear before the right. The sound is also quieter at the right ear because the head is between the sound and the ear – the head shadow effect. These two factors combine to allow localization in this plane (Seeley, Stephens et al. 2004; Moller 2006; Martin and Clark 2011).

The external auditory canal is evolved in a mechanism to protect the ear. The wax and hairs localized in the auditory canal have some protective properties by trapping air-bone particles before they get too deep into the canal. The wax also has some mild antibacterial properties and helps with moisture regulation in the canal: fresh wax is moisture giving and old wax absorbs water. The outer ear is thus a self-cleaning system. Furthermore, the external ear canal has one other function. As a cylinder closed at one end, it has a resonant frequency whose wavelength is four times the length of the canal or approximately 100 mm. This equates to a sound of approximately 3 kHz, and the canal contributes to some amplification of sounds around this frequency (Seeley, Stephens et al. 2004; Moller 2006; Martin and Clark 2011).

Sound waves travel relatively slowly in air, 332 m/s, and a significant time interval may elapse between the time a sound wave reaches one ear and the time that it reaches the

other. The brain can interpret this interval to determine the direction from which a sound is coming (Seeley, Stephens et al. 2004; Moller 2006; Martin and Clark 2011).

b) Middle Ear

Sound waves strike the tympanic membrane and cause it to vibrate. This vibration causes vibration of the three ossicles of the middle ear, and by this mechanical linkage, vibration is transferred to the oval window. More force is required to cause vibration in a liquid like the perylimph of the inner ear than is required in air; thus, the vibrations reaching the perylimph must be amplified as they cross the middle ear. The footplate of the stapes and its annular ligament, which occupy the oval window, are much smaller than the tympanic membrane. Because of this size different, the mechanical force of the vibration is amplified about 20-fold as it passes from the tympanic membrane, through the ossicles, and to the oval window (Seeley, Stephens et al. 2004; Moller 2006; Martin and Clark 2011).

The two covered openings, which are in the medial to the tympanic membrane, the round and oval windows provide air passages from the middle ear. One passage opens into the mastoid air cells in the mastoid process of the temporal bone. The other passageway, the auditory, or Eustachian tube, opens into the pharynx and equalizes air pressure between the outside air and the middle ear cavity. Unequal pressure between the middle ear and the outside environment can distort the eardrum, dampen its vibrations, and make hearing difficult. Distortion of the eardrum, which occurs under these conditions, also stimulates pain fibers associated with it. Because of this distortion, when a person changes altitude, sounds seem muffled, and the eardrum may become painful. These symptoms can be relieved by opening the auditory tube to allow air to pass through to equalize the air pressure. Swallowing, yawning, chewing, and holding the nose and mouth shut while gently trying to force air out of the lungs are methods used to open the auditory tube (Seeley, Stephens et al. 2004; Moller 2006; Martin and Clark 2011).

Foreign objects thrust into the ear, pressure, or infections of the middle ear can rupture the tympanic membrane. Sufficient differential pressure between the middle ear and the outside air can also cause rupture of the tympanic membrane. This can occur in flyers, drives, or individuals who are hit on the side of the head by an open hand (Moller 2006).

Two small skeletal muscles are attached to the ear ossicles and reflexively dampen excessively loud sounds. This sound attenuation reflex protects the delicate ear structures from damage by loud noises. The tensor tympani muscle is attached to the malleus and is innervated by the trigeminal nerve. The stapedius muscle is attached to the stapes and is supplied by the facial nerve. The sound attenuation reflex responds

most effectively to low-frequency sounds and can reduce by a factor of 100 minutes, in response to the prolonged noise (Martin and Clark 2011).

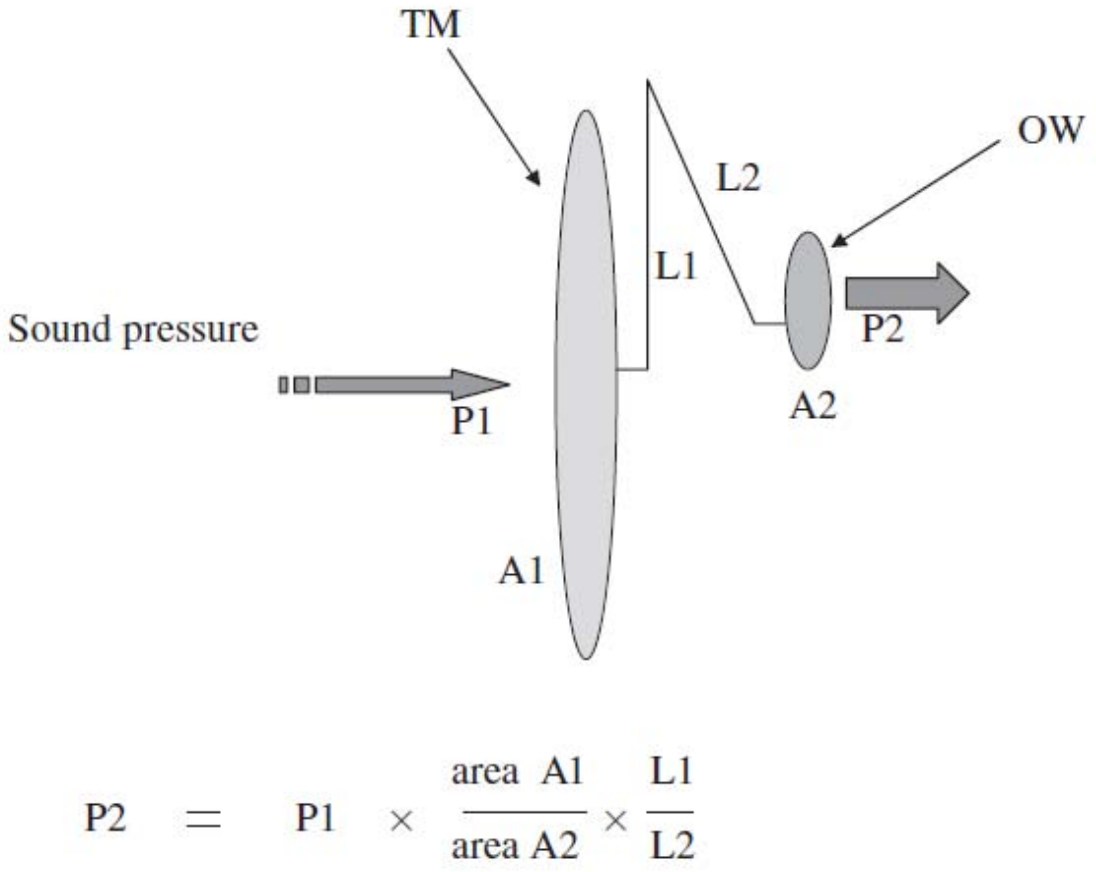


Figure 2.11: A diagram to illustrate the impedance matching mechanism (or transformer mechanism) of the middle ear: P1 = pressure at the tympanic membrane; P2 = pressure at the oval window (OW); A1 = area of the tympanic membrane; A2 = area of the oval window; L1 = manubrium lever of the malleus; L2 = long process of the incus lever (from (Irwin 2006)).

c) Inner Ear

As the stapes vibrates, it produces waves in the perilymph of the scala vestibule. Vibrations of the perilymph are transmitted through the thin vestibular membrane and cause simultaneous vibrations of the endolymph. The mechanical effect is as though the perilymph and endolymph were a single fluid. Vibration of the endolymph causes distortion on the basilar membrane. Waves in the perilymph of the scala vestibule are transmitted through the helicotrema and into the scala tympani. Because the helicotrema is very small, this transmitted vibration is probably of little consequence. Distortions of the basilar membrane, together with weaker waves coming through the helicotrema,

cause waves in the scala tympani perilymph and ultimately result in vibration of the membrane of the round window. Vibration of the round window membrane is important to hearing because it acts as a mechanical release for waves from within the cochlea. If this window were solid, it would reflect the waves, which would interfere with and dampen later waves. The round window also allows relief of pressure in the perilymph because fluid is not compressible, thereby preventing compression damage to the spiral organ (Irwin 2006; Martin and Clark 2011).

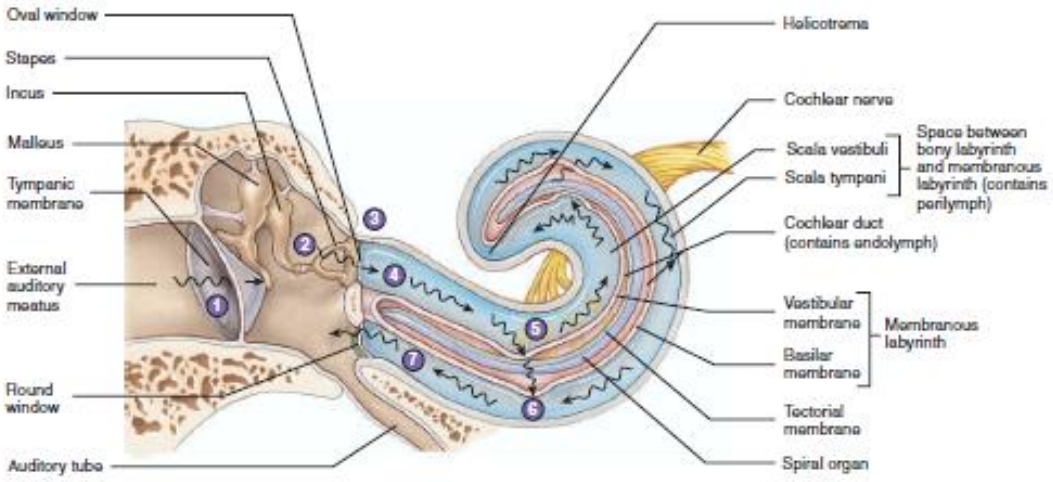


Figure 2.12: Effects of sound waves on cochlear structures (from (Seeley, Stephens et al. 2004)).

The distortion of the basilar membrane is most important to hearing. As this membrane distorts, the hair cells resting on the basilar membrane move relative to the tectorial membrane, which remains stationary. The hair cell microvilli, which are embedded in the tectorial membrane, become bent, causing depolarization of the hair cells. The hair cells then induce action potentials in the cochlear neurons that synapse on the hair cells, apparently by direct electrical excitation through electrical synapse, rather than by neurotransmitters (Seeley, Stephens et al. 2004).

The hairs of the hair cells are bathed in endolymph. Because of the difference in the potassium ion concentrations between the perilymph and endoplymph, an approximately 80 mV potential exists across the vestibular membrane between the two fluids. This is called the endocochlear potential. Because the hair cells are surrounded by endolymph, the hairs have a greater electric potential than if they were surrounded by perilymph. It is believed that this potential difference makes the hair cells much more sensitive to slight movement than they would be if surrounded by perilymph (Irwin 2006; Moller 2006).

The part of the basilar membrane that distorts as a result of endolymph vibration depends on the pitch of the sound that created the vibration and, as a result, on the vibration frequency within the endolymph. The width of the basilar membrane and the length and diameter of the collagen fibers stretching across the membrane at each level along the cochlear duct determine the location of the optimum amount of basilar membrane vibration produced by a given pitch. Higher-pitched tones cause optimal vibration near the base, and lower-pitched tones cause optimal vibration near the apex of the basilar membrane. As the basilar membrane vibrates, hair cells along a larger part of the basilar membrane are stimulated. In areas of minimum vibration, the amount of stimulation may not reach the threshold. In other areas, a low frequency of afferent action potentials may be transmitted, whereas in the optimally vibrating regions of the basilar membrane, a high frequency of action potentials is initiated (Martin and Clark 2011).

Afferent action potentials conducted by cochlear nerve fibers from all along the spiral organ terminate in the superior olivary nucleus in the medulla oblongata. These action potentials are compared to one another, and the strongest action potential, corresponding to the area of maximum basilar membrane vibration, is taken as standard. Efferent action potentials then are sent from the superior olivary nucleus back to the spiral organ to all regions where the maximum vibration did not occur. These action potentials inhibit the hair cells from initiating additional action potentials in the sensory neurons. Thus, only action potentials from regions of maximum vibration are received by the cortex, where they become consciously perceived (Seeley, Stephens et al. 2004). By this process, tones are localized along the cochlea. As a result of this localization, neurons along a given portion of the cochlea send action potentials only to the cerebral cortex in response to specific pitches. Action potentials near the base of the basilar membrane stimulate neurons in a certain part of the auditory cortex, which interpret the stimulus as a high-pitched sound, whereas action potentials from the apex stimulate a different part of the cortex, which interprets the stimulus as a low-pitched sound (Irwin 2006).

Sound volume, or loudness, is a function of sound wave amplitude. As high-amplitude sound waves reach the ear, the perilymph, endolymph, and basilar membrane vibrate more intensely, and the hair cells are stimulated more intensely. As a result of the increased stimulation, more hair cells send action potentials at a higher frequency to the cerebral cortex, where this information is perceived as a greater sound volume (Moller 2006).

Prolonged or frequent exposure to excessively loud noises can cause degeneration of the spiral organ at the base of the cochlea, resulting in high-frequency deafness. The actual amount of damage can vary greatly from person to person. High-frequency loss can cause a person to miss hearing consonants in a noisy setting. Loud music amplified to 120 db, can impair hearing. The defects may not be detectable on routine diagnosis, but they include decreased sensitivity to sound in specific narrow frequency ranges and a

decreased ability to discriminate between two pitches. However, loud music is not as harmful as the sound of a nearby gunshot, which is a sudden sound occurring at 140 db. The sound is too sudden for the attenuation reflex to protect the inner ear structures, and the intensity is strong enough to cause auditory damage. In fact, gunshot noise is the most common recreational cause of serious hearing loss (Seeley, Stephens et al. 2004; Irwin 2006; Moller 2006; Martin and Clark 2011).

2.3.2 Balance

The organs of balance are divided structurally and functionally into two parts. The first, the static labyrinth, consists of the utricle and saccule of the vestibule and is primarily involved in evaluating the position of the head relative to gravity, although the system also responds to linear acceleration or deceleration, such as when a person is in a car that is increasing or decreasing speed. The second, the kinetic labyrinth, is associated with the semicircular canals and is involved in evaluating movements of the head (Seeley, Stephens et al. 2004).

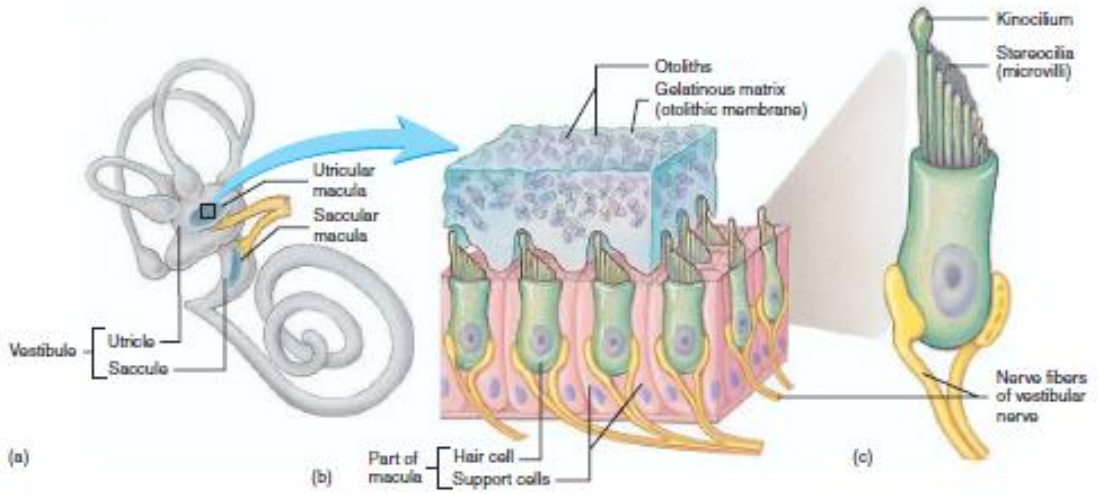


Figure 2.13: Structure of the Macula: Vestibule showing the location of the utricular and saccular maculae (a). Enlargement of the utricular macula, showing hair cells and otoliths in the macula (b). An enlarged hair cell, showing the kinocilium and stereocilia (c) (from (Seeley, Stephens et al. 2004)).

The macula, which exists in a simple cuboide epithelium of the utricle and saccula, is oriented parallel to the base of the skull in the utricle, and the macula of the saccule is perpendicular to the base of the skull (Moller 2006).

The maculae resemble the spiral organ and consist of columnar supporting cells and hair cells. The "hairs" of these cells, which consist of numerous microvilli, called stereocilia, and one cilium, called a kinocilium, are embedded in a gelatinous mass weighted by the presence of otoliths composed of protein and calcium carbonate, Figure 2.13. The gelatinous mass moves in response to gravity, bending the hair cells and initiating action potentials in the associated neurons. Deflection of the hairs toward the kinocilium results in depolarization of the hair cell, whereas deflection of the hairs away from the kinocilium results in hyperpolarization of the hair cell. If the head is tipped, otoliths move in response to gravity and stimulate at low level by the presence of the otolith-weighted covering of the macula; but as this covering moves in response to gravity, the pattern and intensity of hair cell stimulation changes, Figure 2.14. This pattern of stimulation and the subsequent pattern of action potentials from the numerous hair cells of the maculae can be translated by the brain into specific information regarding head position or acceleration. Much of this information is not perceived consciously, but is dealt with subconsciously. The body responds by making subtle tone adjustments in muscles of the back and neck, which are intended to restore the head to its proper neutral, balanced position (Seeley, Stephens et al. 2004; Martin and Clark 2011).

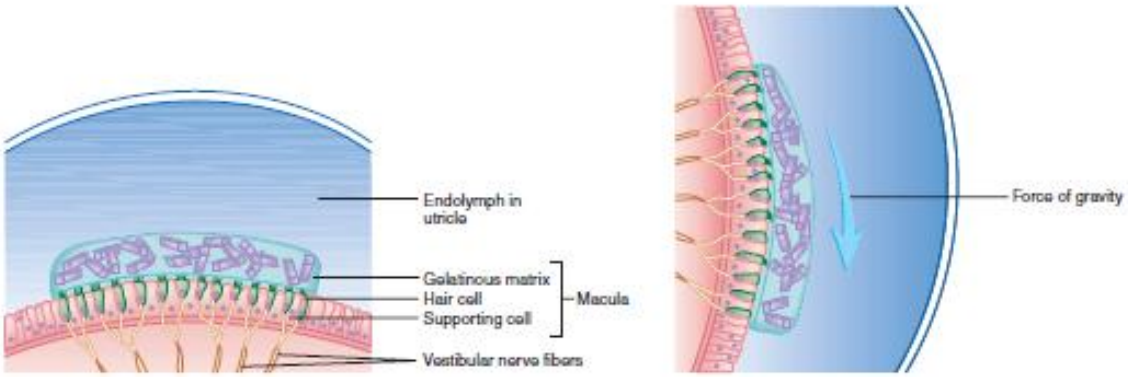


Figure 2.14: Function of the Vestibule in Maintaining Balance (from (Seeley, Stephens et al. 2004)).

The kinetic labyrinth consists of three semicircular canals placed at nearly right angles to one another, one lying nearly in the transverse plane, one in the coronal plane, and one in the sagittal plane. The arrangement of the semicircular canals enables a person to detect movement in all directions. The base of each semicircular canal is expanded into an ampulla. Within each ampulla, the epithelium is specialized to form a crista ampullaris. This specialized sensoty epithelium is structurally and functionally very similar to that of the epithelium maculae, Figure 2.15. Each crista consists of a ridge or crest of epithelium with a curved gelatinous mass, the cupula, suspended over the crest. The hairlike processes of the crista hair cells, similar to those in the maculae, are embedded in the cupula. The cupula contains no otoliths and therefore does not respond to gravitational pull. Instead, the cupula is a float that is displaced by fluid movements within the semicircular canals. Endolymph movement within each semicircular canal

moves the cupula, bends the hairs, and initiates action potentials (Seeley, Stephens et al. 2004; Irwin 2006).

As the head begins to move in a given direction, the endolymph does not move at the same rate as the semicircular canals. This difference causes displacement of the cupula in a direction opposite to that of the movement of the head, resulting in a relative movement between the cupula and the endolymph. As movement continues, the fluid of the semicircular canals begins to move and "catches up" with the cupula and stimulation is stopped. As movement of the head ceases, the endolymph continues to move because of its momentum, causing displacement of the cupula in the same direction as the head had been moving. Displacement of the cupula is most intense when the rate of head movement changes, and this system detects changes in the rate of movement rather than movement alone. As with the static labyrinth, the information obtained by the brain from the kinetic labyrinth is largely subconscious (Seeley, Stephens et al. 2004).

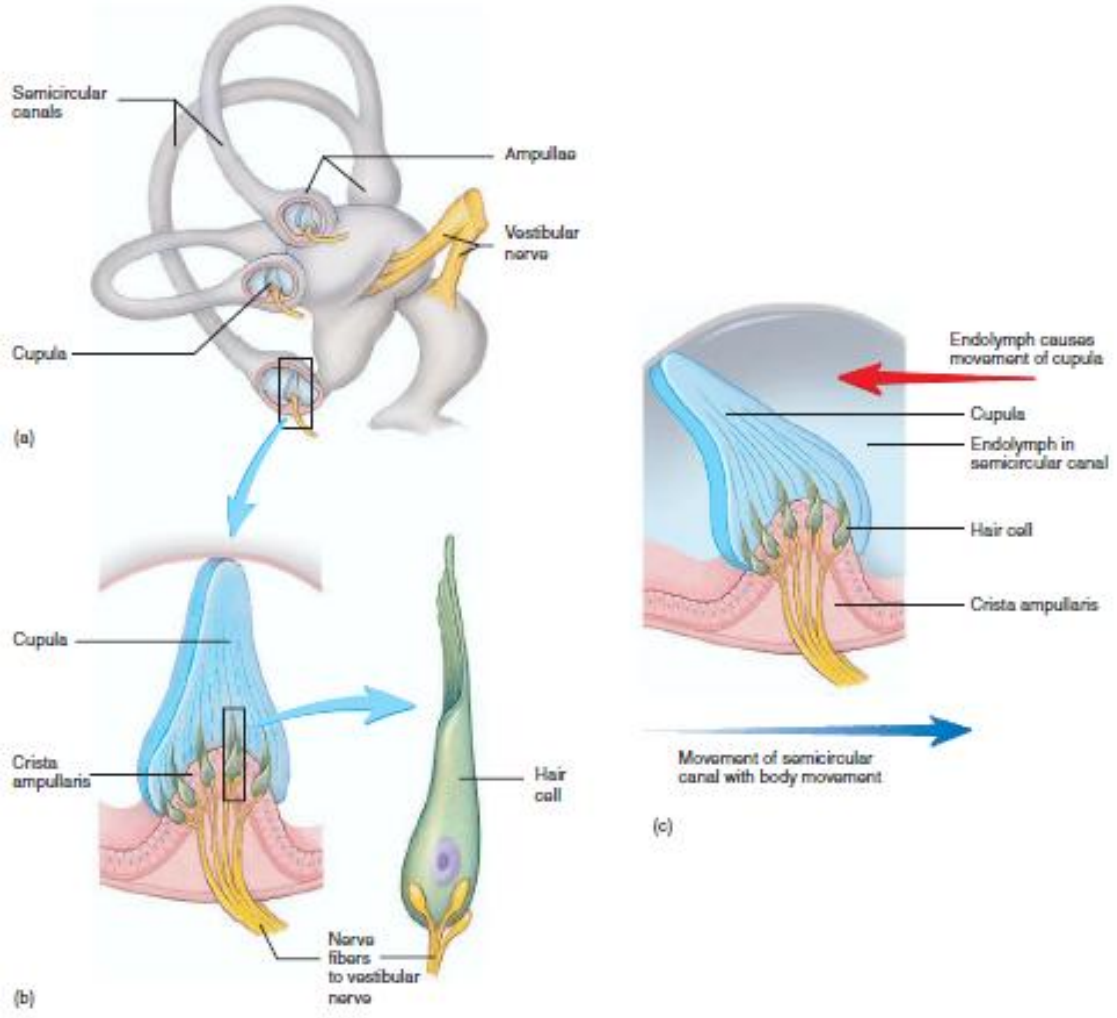


Figure 2.15: Semicircular canals and its function: Semicircular canals showing localization of the crista ampullaris in the ampullae of the semicircular canals (a). Enlargement of the crista

ampullaris, showing the cupula and hair cell (b). The crista ampullaris responds to fluid movements within the semicircular canals (c) (from (Seeley, Stephens et al. 2004)).

2.4 Clinical Focus of Ear Disorders

As people age the number of hair cells decreases in the cochlea. This decline does not occur equally in both ears. As a result, because direction is determined by comparing sounds coming into each ear, elderly people may experience a decreased ability to localize the origin of certain sounds. In some people, this may lead to a general sense of disorientation. In addition, CNS defects in the auditory pathways can result in difficulty understanding sounds with echoes or background noise. Such deficit makes it difficult for elderly people to understand rapid or broken speech (Seeley, Stephens et al. 2004).

Also with age, the number of hair cells in the saccule, utricle, and ampullae decrease. The number of otoliths also declines. As a result, elderly people experience a decreased sensitivity to gravity, acceleration, and rotation. Due of these decreases elderly people experiences dizziness and vertigo. They often feel that they cannot maintain posture and are prone to falling (Seeley, Stephens et al. 2004; Moller 2006). Some examples of ear disorders are: otosclerosis, tinnitus, motion sickness, otitis media and earache.

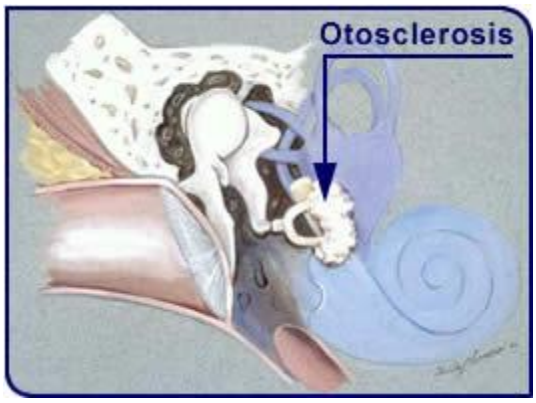


Figure 2.16: Image of otosclerosis occurring between the stapes and the oval window of the cochlea; Notice the buildup of calcium (from (Cannon 2007)).

Otosclerosis is an ear disorder in which spongy bone grows over the oval window and immobilizes the stapes, Figure 2.16. This disorder can be surgically corrected by breaking away the bony growth and immobilizes stapes. During surgery, the stapes is replaced by a small rod connected by a fat pad or a synthetic membrane to the oval

window at one end and to the incus at the other end (Seeley, Stephens et al. 2004; Moller 2006).

Another disorder, tinnitus consists of noises such as ringing, clicking, whistling, or booming in the ears. These noises may occur as a result of disorders in the middle or inner ear or along the central neuronal pathways (Seeley, Stephens et al. 2004; Moller 2006).

Motion sickness is another disorder that affects the structures of the ear. This consists of nausea, weakness, and other dysfunctions caused by stimulation of the semicircular canals during motion, such as in a boat, automobile, airplane, swing, or amusement park ride. It may progress to vomiting and incapacitation. Antiemetics such as anticholinergic or antihistamine medications can be taken to counter the nausea and vomiting associated with motion sickness. Scopolamine is an anticholinergic drug that reduces the excitability of vestibular receptors. Cyclizine, dimenhydrinate and diphenhydramine are antihistamines that affect the neural pathways from the vestibule. Scopolamine can be administered transdermally in the form of a patch placed on the skin behind the ear (Seeley, Stephens et al. 2004; Moller 2006).

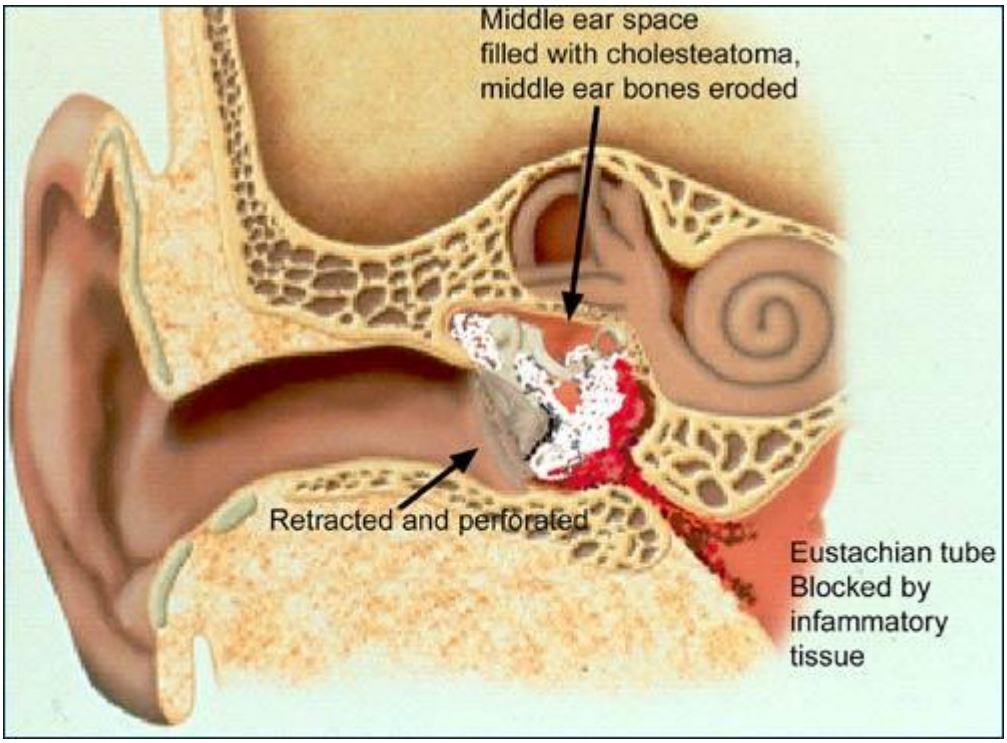


Figure 2.17: Image of otitis media (from (UMMC 2011)).

Infections of the middle ear, known as otitis media, are quite common in young children. These infections usually result from the spread of infection from the mucous

membrane of the pharynx through the auditory tube to the mucouslining of the middle ear. The symptoms of the otitis media, Figure 2.17, consisting of low-grade fever, lethargy, and irritability, are often not easily recognized by the parent as signs of middle ear infection. The infection can also cause a temporary decrease or loss of hearing, because fluid buildup has dampened the tympanic membrane or ossicles (Seeley, Stephens et al. 2004; Moller 2006; Martin and Clark 2011).

Finally, earache can result from otitis media, external otitis (inflammation of the external auditory meatus), dental abscesses, or temporo mandibular joint pain. The Eustachian tube dysfunction occurs when the tube fails to open during swallowing or yawing, which results in a difference between the air pressure inside and outside the middle ear. It causes discomfort in the ear, hearing problems and will usually result in fluid accumulation behind the eardrum (Seeley, Stephens et al. 2004; Moller 2006).

2.5 Summary

The external ear consists of the auricle and external auditory meatus and the auditory tube connects the middle ear to the pharynx. The principal function is to equalize pressure.

The middle ear connects the external and inner ear. In the external auditory meatus, the tympanic membrane is stretched, the malleus, incus, and stapes connect the tympanic membrane to the oval window of the inner ear and the middle ear is connected to the mastoid air cells.

The inner ear has three parts: the semicircular canals, the vestibule, which contains the utricle and saccule, and the cochlea. The cochlea is a spiral-shaped canal within the temporal bone and is divided into three compartments by the vestibular and basilar membranes. The perilymph is in the interior of the scala vestibule and scala tympani, but the cochlear duct contains endolymph and the spiral organ or organ of the corti. The spiral organ consists of hair cells that attach to the tectorial membrane.

The auditory function is characterized by a variety of processes. First, sound waves are funneled by the auricle down the external auditory meatus, causing the tympanic membrane to vibrate. The second step consists in the passing of the tympanic membrane vibrations through the auditory ossicles to the oval window of the inner ear. Then, movement of the stapes in the oval window causes the perilymph, vestibular membrane, and endolymph to vibrate, producing movement of the basilar membrane. Movement of the basilar membrane causes displacement of the hair cells in the spiral organ and the generation of action potentials, which travel along the vestibulocochlear nerve.

Finally, some vestibulecochlear nerve axons synapse in the superior olivary nucleus and efferent neurons from this nucleus project back to the cochlea, where they regulate the

perception of pitch. The round window is responsible for protect the inner ear from pressure buildup and dissipates waves.

The balance can be static or kinetic. In the first case, static balance evaluates the position of the head relative to gravity and detects linear acceleration and deceleration. For this case, the maculae of the utricle and saccule consists of hair cells with hairs embedded in gelatinous mass that contains otoliths. The gelatinous mass is responsible for movement in response to gravity. On the other hand, kinetic balance evaluates movements of the head. When the head moves, endolymph within the semicircular canal moves the cupula. Present in the inner ear are three semicircular canals at right angles to one another and the ampulla of each semicircular canal contains the crista ampullaris, which has hair cells with hairs embedded in a gelatinous mass, the cupula.

Chapter III – Image Segmentation Algorithms

- Introduction;
- Segmentation Algorithms;
- Summary.

3.1 Introduction

Segmenting structures from medical images and reconstructing a compact geometric representation of these structures is difficult due to the sheer size of the datasets and the complexity and variability of the anatomic shapes of interest. Furthermore, the shortcomings typical of sampled data, such as sampling artifacts, spatial aliasing, and noise, may cause the boundaries of structures to be indistinct or disconnected. The challenge is to extract boundary elements belonging to the same structure and integrate these elements into a coherent and consistent model of the structure. Traditional low-level image processing techniques which consider only local information can make incorrect assumptions during this integration process and generate infeasible object boundaries. As a result, these model-free techniques usually require considerable amounts of expert intervention. Furthermore, the subsequent analysis and interpretation of the segmented objects is hindered by the pixel- or voxel-level structure representations generated by most image processing operations (McInerney and Terzopoulos 1996).

This chapter surveys computational medical image segmentation algorithms, which are now used routinely in a multiple of different applications, such as the quantification of tissue volumes, diagnosis, localization of pathology, study of anatomical structure, treatment planning, partial volume correction of functional imaging data, and computer-integrated surgery. Medical image segmentation algorithms are reviewed in this chapter.

3.2 Segmentation Algorithms

Image segmentation is a basic problem in Computational Vision and the key to the procedure from processing to analyzing. Extracting object contours from medical images can help doctors understand diseases and play an important role for better diagnosis and treatment plans (Xiao-Juan and Dan 2010).

Segmentation is made during automated analysis by delineating structures of interest and discriminating them from background tissue, and this separation process is generally effortless and swift for the human visual system. However, the separation process performed by the human visual system can become a considerable challenge in algorithm development (Bankman 2000).

For the analysis of the medical images, segmentation is important for feature extraction, image measurements, and image display. Furthermore, it may be useful to classify image pixels into anatomical regions, such as bones, muscles, and blood vessels, while

in others into pathological regions, such as cancer, tissue deformities, and multiple sclerosis lesions (Bankman 2000).

The segmentation of organs or structures in medical images is still often performed manually. The manual segmentation medical technicians require to sketch the contours slice by slice using pointing devices, such as a mouse or a trackball, and are very time-consuming, besides that the results may change from intra- or inter- observer variability. To solve the manual segmentation disadvantages, modern mathematical and physical techniques have been integrated into computational approaches. These integrations have greatly enhanced the accuracy of the segmentation results (Ma, Tavares et al. 2010).

Medical images can be induced by noise and artifacts. One kind of artifacts is the partial-volume effects, which occur where multiple tissue types contribute to a single pixel, resulting in a blurring of intensity across boundaries. Unfortunately, partial-volume effects are common in medical images, particularly for computerized tomography images and magnetic resonance images (Pham, Xu et al. 2000). Thus, it becomes essential to incorporate physical and mathematical techniques to increase the accuracy of segmentation results. The preprocessing of the input images is usually the first event performed to reduce the image noise and increase the contrast of the structures of interest. This stage is really necessary because an accurate interpretation of the medical images may become difficult if noise levels are relatively high, and the enhancement improves the quality of the input image and thus facilitates diagnosis. Segmentation is often included in the analysis stage and can be divided into three classes: thresholding, clustering and deformable models (Ma, Tavares et al. 2010). Following, the common segmentation algorithms are introduced.

3.2.1 Algorithms based on Tresholding

Thresholding is a common region segmentation method. In this technique is nominated a threshold value and the image is divided into groups of pixels that have values less than the threshold value, and groups of pixels with values greater or equal to the threshold value (Bankman 2000). Thresholding is a simple yet often effective means for obtaining a segmentation of images in which different structures have contrasting intensities or other quantifiable features, and it is frequently performed interactively, based on the operator's visual assessment of the resulting segmentation (Pham, Xu et al. 2000).

There are several thresholding techniques; some of them, are based on the image histogram and others are based in local properties or local gradient. The global thresholding is the most intuitive approach, and is called global as only one threshold

value is selected for the entire image, based on the image histogram. When the threshold value depends on local properties of some image regions, it is called local. On the other hand, thresholding is called dynamic or adaptive if the local thresholds are selected independently for each pixel (Bankman 2000; Xiao-Juan and Dan 2010).

Thresholding is often used as an initial step in a sequence of image-processing operations and its main limitations are that, in its simplest form, only two classes are generated and, typically, it does not take into account the spatial characteristics of the input image. In addition, this causes it to be sensitive to noise and intensity inhomogeneity. These artifacts corrupt the histogram of the image and making separation more difficult (Pham, Xu et al. 2000).

At the present, there are several solutions to selecting the threshold value, such as: histogram trough method, Otsu method, maximum entropy method, vector retention method and grade statistic method. However, Otsu method proposed in 1979 is the most popular threshold-based method. This method is deduced by least square (LS) technique based on gray histogram. As is well known, this method not only has the best threshold value in the statistical sense, but also it is the most stable method in the image threshold segmentation at present. Moreover, its basic principle is to split the image's pixels into two classes, and confirms the best threshold value through the variance maximum value between the two classes.

Supposed that $G = [0, L-1]$ is the range of a grayscale image $f(x, y)$ and P_i is the probability of every grayscale, and the threshold value t has splitted the image in two classes which are $C_0 = [0, t]$ and $C_1 = [t+1, L-1]$. The two classes probability are $\alpha_0 = \sum_{i=0}^{t} P_i$ and $\alpha_1 = 1 - \alpha_0$, respectively. The average gray value of the two classes are $\mu_0 = \sum_{i=0}^{t} \frac{iP_i}{\alpha_0} = \frac{\mu_t}{\alpha_0}$ and $\mu_1 = \sum_{i=t+1}^{L-1} \frac{iP_i}{\alpha_1} = \frac{\mu - \mu_t}{1-\alpha_0}$, respectively, there in $\mu = \sum_{i=0}^{L-1} iP_i$, $\mu_t = \sum_{i=0}^{t} iP_i$. The criterion function has been defined as variance between the two classes, expressed as:

$$\eta^2(t) = \alpha_0(\mu_0 - \mu)^2 + \alpha_1(\mu_1 - \mu)^2 = \alpha_0\alpha_1(\mu_0 - \mu_1)^2. \quad (3.1)$$

Calculating Eq.(3.1), the maximum t, which is the threshold value, can be obtained (Fang, Yue et al. 2009).

Local thresholding algorithms can be further classified as edge-based, region-based and hybrid. Edge-based algorithms are associated to the edge information, and the structures are described by edge points. Edges are significant local changes of intensity in an image, and they typically occur on the boundary between two different regions in an image. The fundamental goals of edge-detection are two: produce a line drawing of a scene from an image of that scene and extract important features from the edges of an image (e.g. corners, lines, curves). The wavelet transform, Canny edge detection, Sobel

edge detection and Laplacian edge detection are based on threshold values and use the edge information. Furthermore, Laplacian edge detection and Canny edge detection are algorithms that find edge pixels and eliminate the noise influence. Laplacian edge detection uses the second derivation information of the image intensity. On the other hand, Canny edge detection uses the gradient magnitude to find edge pixels and suppresses them through non-maximal suppression and hysteresis thresholding (Ma, Tavares et al. 2010).

The gradient operator is an important filter that is used by Canny, Sobel and Laplacian detector, it produces a vector at each point, whose direction gives the maximum change of the function at that point, and whose magnitude gives the magnitude of this maximum change. Furthermore, the gradient operator is often computed by convolving two windows with an image as:

$$G_x(i,j) = mask_x * f(i,j) \qquad G_y(i,j) = mask_y * f(i,j), \tag{3.2}$$

where $f(i,j)$ indicates some neighborhood of pixel (i,j), and $*$ denotes convolution (Kanopoulos, Vasanthavada et al. 1988).

The wavelet representation, mentioned above, corresponds to a decomposition of the image into a set of independent frequency bands, and wavelet representation is very useful in edge detection. Furthermore, Akhtar and Ali (2008) worked with wavelet representation and exploited three spatial orientations: horizontal, vertical and diagonal orientations. Akhtar and Ali (2008) proposed three stages: edge detection, linking framework using wavelets and image fusion. In the wavelet domain, the detection and linking of edges found very natural and straightforward operation (Akhtar and Ali 2008). The decomposition-stage is the first step. In this first step, the input image is decomposed into four quarter-size output images with different information, Eqs. 3.3 and 3.4, Figure 3.1. The second step is the recomposition-stage that is characterized by recomposition of horizontal and vertical details, Eq. 3.5. In the final step, the recomposed images are decomposed again, as in the first step, to obtain wavelet coefficients, and then these resulting images are fused using Eq. 3.6.

$$W_\varphi(j_0,m,n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f \varphi_{j_0,m,n} \tag{3.3}$$

$$W_\psi^i(j,m,n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f {\psi^i}_{j,m,n} \tag{3.4}$$

$$f = \frac{1}{\sqrt{MN}} \sum_m \sum_n \left\{ W_\varphi^{j_0} \varphi_{j_0} + \sum_i \sum_{j=j_0}^{\infty} W_{\psi^i}^{j} \psi_j^i \right\} \tag{3.5}$$

$$W_\varphi = \begin{cases} W_{\psi^H}, & \left|W_{\psi^H}\right| \geq \left|W_{\psi^V}\right| \\ W_{\psi}{}^{V}, & otherwise \end{cases} \tag{3.6}$$

where $\varphi(x,y)$ is a two dimensional scaling function and in 2D is required three 2D wavelets: $\psi^H(x,y)$, $\psi^V(x,y)$ and $\psi^D(x,y)$ (horizontal (H), vertical (V) and Diagonal (D) details). The input image is represented by $f = f(x,y)$ and has a size of $M \times N$ pixels. So, j, m, n, M and N are a set of integers, while $i = \{H, V, D\}$, j_0 is an arbitrary starting scale (usually, equal to 1) and the coefficients W_φ define an approximation of f at scale j_0 (Akhtar and Ali 2008).

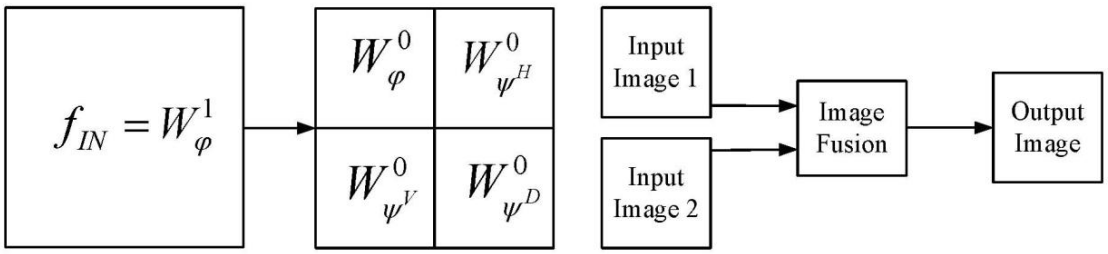


Figure 3.1: Wavelet decomposition (left) and image fusion process (right) (from (Akhtar and Ali 2008)).

Another edge detector is the Canny edge detector, which presents good detection, good localization and single response to an edge, so it is usually considered optimal and popular. A typical implementation of the Canny edge detector follows the following steps: smooth the given image, $f_s(i,j)$, with an appropriate Gaussian, $G(i,j)$, Figure 3.2, to reduce desired amount of image details and noise; determine the gradient magnitude, $m(i,j)$, and gradient direction, $\theta(i,j)$, at each pixel; if the gradient magnitude, Eq. 3.9, at a pixel is larger than those at its two neighbors in the gradient direction, Eq. 3.10, mark the pixel as a major edge and if the gradient magnitude at the pixel is larger than those pixels adjacent to it in any direction mark the pixel as a minor edge; partition the minor edge contours at the branch points, remove all branches that do not contain a major edge, then rename as a major edges the portions of minor edge contours that are delimited by major edges; combine newly obtained major edges with previously obtained major edges and finally, remove from among the combined major edges those that are sufficiently weak by hysteresis thresholding (Ding and Goshtasby 2001).

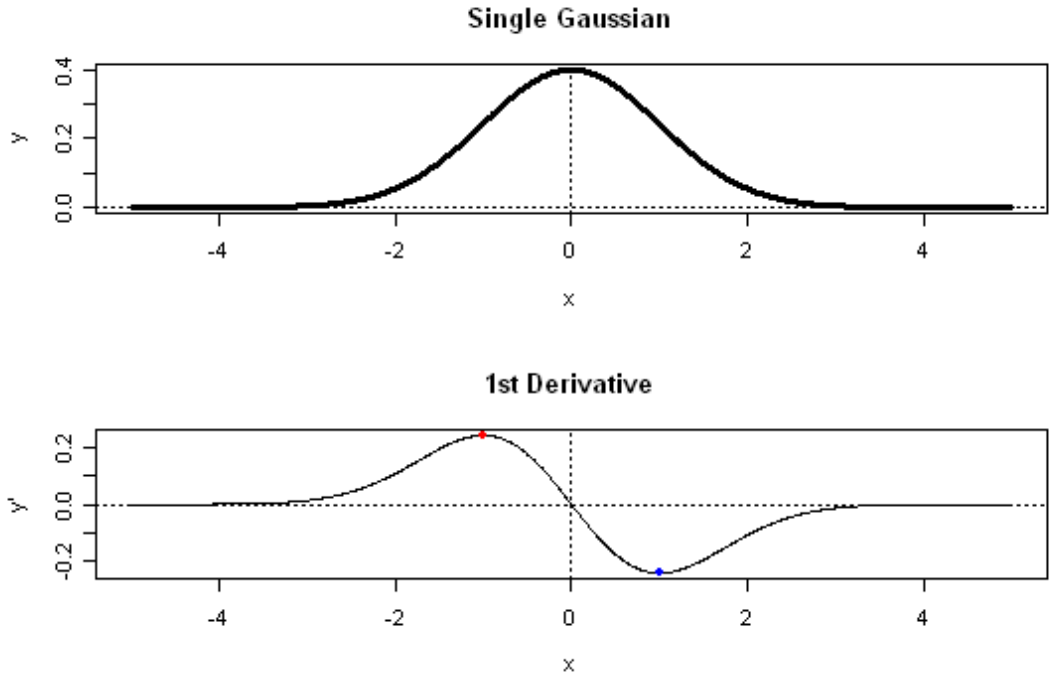


Figure 3.2: Single gaussian with its first derivate (from (Glynn 2007)).

$$f_s(i,j) = G(i,j) * f(i,j) \tag{3.7}$$

$$\nabla f_s = \left[\frac{\partial f_s}{\partial f_x}, \frac{\partial f_s}{\partial f_y}\right]^t \tag{3.8}$$

$$m(i,j) = \sqrt{\left(\frac{\partial f_s}{\partial f_x}\right)^2 + \left(\frac{\partial f_s}{\partial f_y}\right)^2} \tag{3.9}$$

$$\theta(i,j) = \arctan\left(\frac{\frac{\partial f_s}{\partial f_y}}{\frac{\partial f_s}{\partial f_x}}\right) \tag{3.10}$$

The Sobel operator as image edge detection filter is characterized by convolutions between the input image, $f(i,j)$, and four 3×3 masks weighted. These convolutions measure the differences in intensity along the horizontal, vertical, and left and right

diagonal directions. These four measurements, M_H, M_V, M_{DL} and M_{DR}, Eqs. 3.11, 3.12, 3.13 and 3.14, are then combined to estimate edge magnitude and direction.

$$M_H = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * f(i,j) \tag{3.11}$$

$$M_V = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * f(i,j) \tag{3.12}$$

$$M_{DL} = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ 2 & -1 & 0 \end{bmatrix} * f(i,j) \tag{3.13}$$

$$M_{DR} = \begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} * f(i,j) \tag{3.14}$$

The gradient magnitude, Eq. 3.15, is estimated as given by Eq. 3.16,

$$Mag = \left[{M_H}^2 + {M_V}^2\right]^{1/2} \tag{3.15}$$

$$Mag = max[|M_H|, |M_V|, |M_{DR}|, |M_{DL}|] + K[M_\perp] \tag{3.16}$$

and $M_\perp$ is the measure in the direction perpendicular to the selected maximum measure. The other hand, the gradient direction is given by the equation:

$$\theta = \arctan\left[\frac{M_V}{M_H}\right] \tag{3.17}$$

and the gradient direction is given from a template of eight angle directions based on the selected maximum measure and its corresponding sign. For this edge detection filter, the best value for *K* is determined as the value that minimizes the magnitude error, which can be computed using error analysis for Euclidean norm approximations (Kanopoulos, Vasanthavada et al. 1988).

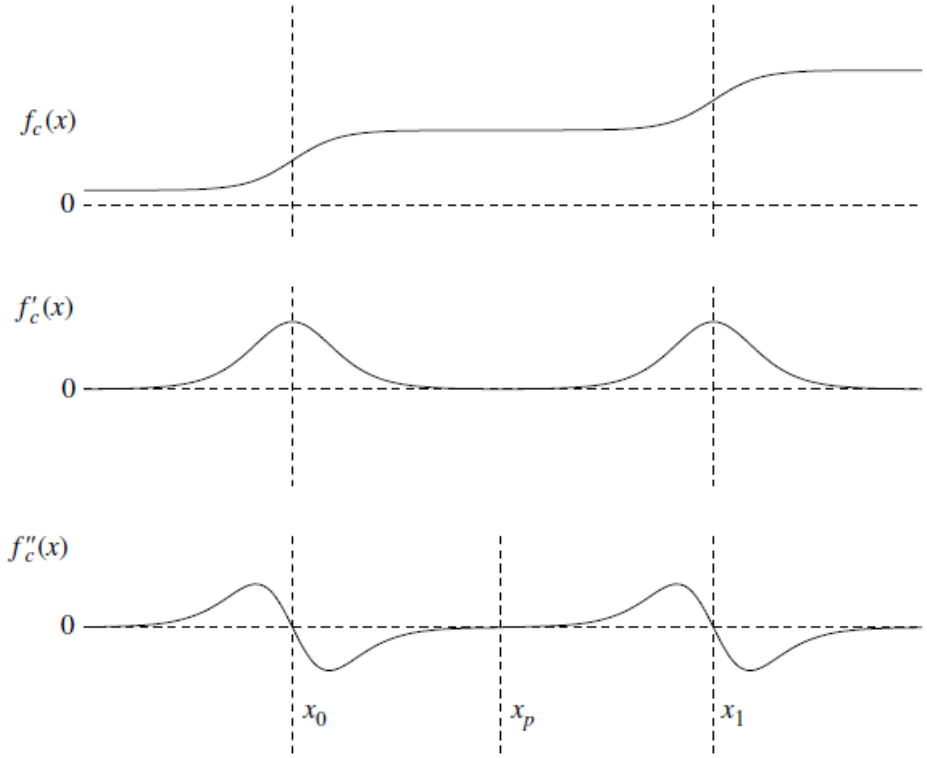


Figure 3.3: The zero-crossing of the second derivate of an edge (from (Mlsna and Rodríguez 2009)).

Edge points also can be detected by finding the zero-crossing of the second derivate and the Laplacian operation corresponds to the second derivate, Figure 3.3. Therefore, the derivate operator acts as a high pass filter and edge detectors based on it are sensitive to noise (Mlsna and Rodríguez 2009). The Laplacian is defined as:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{3.18}$$

$$\frac{\partial^2 f}{\partial x^2} = f(i, j+1) - 2f(i,j) + f(i, j-1) \tag{3.19}$$

$$\frac{\partial^2 f}{\partial^2 y^2} = f(i+1, j) - 2f(i,j) + f(i-1, j) \tag{3.20}$$

$$\nabla^2 f = -4f(i,j) + f(i,j+1) + f(i,j-1) + f(i+1,j) + f(i-1,j) \quad (3.21)$$

and the Laplacian can be implemented using the mask shown in Figure 3.4.

Laplacian is an isotropic operator, and its implementation has low computational cost. Furthermore, this operator does not provide information about edge direction and Laplacian-based edge detection has the nice property that it produces edge of zero thickness, making edge-thinning steps unnecessary. This is because the zero-crossings themselves define the edge location (Mlsna and Rodríguez 2009).

0	1	0
1	-4	1
0	1	0

Figure 3.4: Laplacian mask (3x3).

An edge detector based only on the zero crossing of the continuous Laplacian produces closed edge contours if the input image meets certain smoothness constrains. The contours are closed because edge strength is not considered, so even the slightest, most gradual intensity transition produces a zero crossing. In effect, the zero-crossing contours define the boundaries that separate regions of nearly constant intensity in the original image. The second derivate zero-crossing occurs at the local extreme of the first derivate, but many zero crossings are not local maxima of the gradient magnitude. Some local minima of the gradient magnitude can be largely eliminated by appropriately thresholding the edge strength (Mlsna and Rodríguez 2009).

Noise presents a problem for the Laplacian edge detector in several ways. First, noise produces many false edge contours, because it introduces variation to the constant-intensity regions in the noise-free energy. Second, noise alters the locations of the zero-crossing points, producing location errors along the edge contours. The problem of noise-induced false edges can be addressed by applying an additional test to the zero-crossing points. Only the zero-crossing satisfies this new criterion and are considered edge points. One commonly used technique classifies a zero-crossing as an edge point, if the local gray level variance exceeds a threshold amount. Another method is to select

the strong edges by thresholding the gradient magnitude or the slope of the Laplacian output at the zero-crossing. These techniques serve to reject zero-crossing points that are more likely caused by noise than by a real edge in the original scene. Of course, thresholding the zero-crossings in this manner tends to break up the closed contours (Mlsna and Rodríguez 2009).

Marr and Hildreth advocated the need for an operator that can be tuned to detect edges at a particular scale. Their method is based on filtering the image with a Gaussian kernel selected for a particular edge scale:

$$G(i,j) = e^{-\frac{i^2+j^2}{2\sigma^2}} \tag{3.22}$$

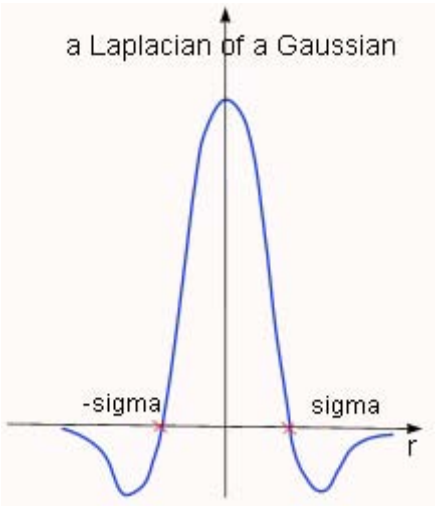


Figure 3.5: Characteristic curve of log filter (from (McAullife 2010)).

The Gaussian smoothing operation serves to band-limit the image to a small range of frequencies, reducing the noise sensitivity problem when detecting zero-crossings (Mlsna and Rodríguez 2009). This new operator is called the Laplacian-of-Gaussian (LOG), Figure 3.5, and it can be shown that:

$$\nabla^2[f(i,j) * G(i,j)] = \nabla^2 G(i,j) * f(i,j) \tag{3.23}$$

$$\nabla^2 G(i,j) = \left(\frac{r^2-\sigma^2}{\sigma^4}\right) e^{-r^2/2\sigma^2}, (r^2 = i^2 + j^2) \tag{3.24}$$

The detected boundaries consist of discrete pixels and may be incomplete or discontinuous. To connect the breaks or eliminate the holes it is necessary to apply morphological operations (Ma, Tavares et al. 2010). Dilation and erosion are classic operators for performed mathematical morphological operations. These operators work with the original image and with a structuring element (Bosworth and Acton 2003; Bai, Zhou et al. 2009). Dilation is characterized by production of clear images (maximum operation), it is observed when the resulting images are compared with original images, whereas is verified the opposite for erosion (minimum operation). Thus erosion and dilation are not invertible operations, and both simplify and remove information from the input images. The removed information increased with the size of structuring element. Therefore, the shape and size of this element are important features to obtain efficient results. Furthermore, the realization of several dilations or erosions with small structuring elements is not equivalent to dilation or erosion realized with a big structuring element. One disadvantage of the morphological operations is the high noise sensitive (Bosworth and Acton 2003). The morphological representation supports the basic objectives of the image pre-processing and is an inefficient method for segmenting anatomic structures of medical images (Papari and Petkov 2011).

In medical images, the edges of structures are usually not clearly defined due to noise influence and partial-volume effects (PVE) (Ma, Tavares et al. 2010), which are artifacts that occur where multiple tissue types contribute to a single pixel and result in a blurring of intensity across boundaries (Pham, Xu et al. 2000). Therefore, algorithms based on edge detection are seldom used alone but instead as an efficiency pre-processing step for the later segmentation step (Ma, Tavares et al. 2010).

The region-based algorithms are another type of thresholding-based algorithms, and their idea comes from the observation that quantifiable features inside a structure tend to be homogeneous (Ma, Tavares et al. 2010). Therefore, algorithms aim to search for the pixels with similar feature values. Classically, examples of this type are the region growing algorithms and split and merge algorithms (Bankman 2000). Region Growing starts with a pixel or a group of pixels which are called seeds and can be chosen by an operator. Seeds also can be provided by an automatic seed finding procedure. For region growing procedure the next step is to examine the neighboring pixels, so they are analyzed one at a time and added to the growing region, if they are sufficiently similar based on a uniformity test. The procedure continues until no more pixels can be added. The result is then represented by all pixels that have been accepted during the growing procedure. The uniformity test can be compared with the difference between the pixel intensity value and the mean intensity value, and when the difference is less than a predefined value, the pixel is included in the region, otherwise, is defined as an edge pixel. The results of region growing depend strongly on the selection of the homogeneity criterion. If this criterion is not properly chosen, the regions leak out into adjoining areas, or merge with regions that do not belong to the object of interest (Bankman 2000; Pham, Xu et al. 2000). Region growing is seldom used alone, but

usually within a set of image-processing operations, particularly for the delineation of small, simple structures such as tumors and lesions. The primary disadvantage of region growing is that it requires manual interaction to obtain the seed point. Thus, for each region that needs to be extracted, a seed must be planted (Pham, Xu et al. 2000). Another problem of region growing is that different starting points may not grow into identical regions. Otherwise, the advantage of region growing is that it is capable of correctly segmenting regions that have the same properties and are spatially separated. Another advantage is that it generates connected regions. However, region growing can also be sensitive to noise, caused extracted regions to have holes or even become disconnected. Conversely, partial-volume effects can cause separate regions to become connected (Bankman 2000; Pham, Xu et al. 2000).

Instead of region merging, it is possible start with some initial segmentation and subdivide the regions that not satisfy a given uniformity test and this technique is called splitting. A combination of splitting and merging adds together the advantages of both approaches. Furthermore, split-and-merge not requires a seed point (Bankman 2000). The split-and-merge objective is to automatically segment an image into the minimum number of regions that faithfully represent areas of uniformity, producing boundaries with characteristics well matched to the intrinsic image resolution (Manousakas, Undrill et al. 1998).

Finally, hybrid algorithms combine different image cues to complete the segmentation, and a typical example is the watershed algorithm (Ma, Tavares et al. 2010). Watershed segmentation is a technique that utilizes image morphology, is applicable to non-specific image type, and is also unsupervised. The watershed method can be classified into rain falling and water immersion. Some watershed methods use the prior information-based difference function, instead of the more frequently-used gradient function to improve the segmented results and employ the marker images as probes to explore a gradient space of unknown image and thus to determine the best matched object (Chen and Chen 2009). Watershed requires selection of at least one marker (seed point) interior to each object of the image, including the background as a separate object. The markers are chosen by an operator or are provided by an automatic procedure that takes into account the application-specific knowledge of the objects. Once the objects are marked, they can be grown using a morphological watershed transformation. Watershed methods can be thought as a surface where the bright pixels represent mountaintops and the dark pixel valleys. The surface is punctured in some of the valleys, and then slowly submerged into a water bath. The water will pour in each puncture and start to fill the valleys. However, the water from different punctures is not allowed to mix, and therefore the dams need to be built at the points of first contact. These dams of the boundaries of the water basins, and also the boundaries of image objects (Bankman 2000). The advantage of the watershed method is that can segment multiple objects in a single threshold setting. On the other hand, the disadvantage of the watershed method is that the different types of images need different thresholds, and if

the thresholds are not set correctly, then the objects are under-segmented or over-segmented. Furthermore, slight changes in the threshold can significantly alter the segmentation results (Chen and Chen 2009).

Threshold-based algorithms are effective when the interested structures have distinctive quantifiable features, and they are computationally efficient. However, these algorithms are sensitive to noise and most of them are difficult to apply to multi-channel images. Furthermore, medical images usually present noise and intensity inhomogeneity; thus, segmentation results of threshold-based algorithms are far from satisfactory and are rarely used alone.

3.2.2 Algorithms based on Clustering

In medical images, techniques from pattern recognition fields can be used to perform the segmentation, and the clustering techniques are the most popular ones for medical image segmentation (Ma, Tavares et al. 2010). The clustering algorithms have the same function as classifier methods and these are pattern recognition techniques that seek to partition a feature space derived from the image using data with known labels (Pham, Xu et al. 2000). In clustering algorithms, the partitioning is based on pairwise relations between individual patterns, rather centralized relations between samples and a few cluster representatives (like centroids, for example) (Matta and Dugelay 2009). Clustering is according to some measures and certain criteria to separate the collection of individuals into several categories, making the degree of similarity among the same individuals as large as possible, but the degree of similarity among different individuals as small as possible.

The basis for many computational applications is the data analysis, and these techniques can be divided into two categories: exploration and confirmation. Clustering is the most commonly used technique among exploratory data analysis techniques, while the classification is the representative technique of confirmatory data analysis techniques. The training data provided with known marks is classification and the training obtains the description mode through learning, then use the model to classify the feature data that is a supervised learning. If the clustering belongs to the unsupervised learning, it main purpose is to divide the unmarked data into meaningful groups. The clustering technique can be divided in the following task phases: data representing, measure, technique and other optional phases. Data representing includes the feature selection and the feature extraction. The feature selection refers to among the collection of all data attributes selecting a subset to represent the data. On the other hand, the feature extraction is the new property generated by the existing data property. The second

phase, measure, uses the distance based or similarity based representation methods. Finally, in the third phase, the clustering results are obtained by a variety of clustering algorithms. Other optional phases are the data abstraction of the cluster results and assessment of the clustering results, for example (Zhu 2010).

Furthermore, the clustering techniques can be divided in three main classes: supervised algorithms, unsupervised algorithms and semi-supervised algorithms (Sutton, Bezdek et al. 2000; Ma, Tavares et al. 2010). The supervised techniques are frequently used and include k-nearest neighbor (kNN), maximum likelihood (ML) algorithms, supervised artificial neuronal networks (ANN), support vector machines (SVM), active shape models (ASM) and active appearance models (AAM) (Ma, Tavares et al. 2010). For perform supervised classification is needed a training set to extract structure information and the key issues of supervised clustering are the tagging data to guide the clustering process, therefore the clustering process will get better effects, and how to access and use marked data (Zhu 2010).

The k-nearest-neighbor is a simple classifier and, through this algorithm, each pixel is classified into the same class as the majority of the k-closest training data (Betanzos, Varela et al. 2000; Pham, Xu et al. 2000). The k-nearest neighbor algorithm needs to be established beforehand, the number of classes in which to divide the set to be classified that, in turn, is necessary to supply with a set of samples, indicating for each one the output cluster to which it belongs. For each one of the nearest k samples, the k-nearest neighbor algorithm determines the value of the correspondence functional and these nearest k samples are known as neighbors. Moreover, the level of confidence can be measured through the obtained results. The following equation is used in order to calculate a pixel's factors of correspondence to clusters being considered:

$$u_i(x) = \frac{\sum_{j=1}^{K} u_{ij}\left(\frac{1}{\left\|x - x_j\right\|^{2/(m-1)}}\right)}{\sum_{j=1}^{K}\left(\frac{1}{\left\|x - x_j\right\|^{2/(m-1)}}\right)} \tag{3.25}$$

where u_{ij} represents the factor of correspondence of the j^{th} sample to the i^{th} class; x_j signifies one of the k samples nearest to the pixel in question; x represents the pixel itself; m is a factor that weights the distance from the pixel to the sample and, finally, $u_i(x)$ represents the degree of correspondence of pixel x to class i. Furthermore, this equation attributes not only a value to the distance of a pixel from all its neighbors, but also to its factor of correspondence to the object cluster for each one of these neighbors. The number of samples, the quality of the samples and the number of neighbors are different parameters that varied in the analysis of this algorithm (Betanzos, Varela et al. 2000).

Maximum likelihood algorithms are another supervised approach, in which the training step of the maximum likelihood algorithm intends to identify the parameters used in the statistical models. In the training of this supervised technique is necessary to identify the parameters that are used in the statistical models. During the processing of maximum likelihood algorithm, the pixel intensities are independent random variables with parameterized probability functions. The evolution of these parameters is realized by maximizing the likelihood function of the mixture model (Ma, Tavares et al. 2010). The iterative computation of maximum likelihood consists of an expectation step followed by a maximization step, and this process is called the EM algorithm (Dempster, Laird et al. 1977). The EM algorithm is originally described in its general form by Dempster (1977) and it is a common method of finding the maximum-likelihood estimation of the parameters of an underlying distribution from a given data set when the data is incomplete or has missing values (Xie and Tsui 2004). Furthermore, the EM process is remarkable in part because of the simplicity and generality of the associated theory. If the underlying complete likelihood estimates are easily computed, then each maximization step of an EM algorithm is likewise simply computed. The complete-data specification $f(\dots|\dots)$ is related to the incomplete-data specification:

$$g(y|\phi) = \int_{\chi(y)} f(x|\phi)dx \tag{3.26}$$

where x is refered as the complete data and y is the observed data.

The EM algorithm is directed at finding a value of ϕ which maximizes $g(y|\phi)$ given an observed y, but it does so by making essential use of the associated family $f(x|\phi)$ (Dempster, Laird et al. 1977). This method does not take advantage of the spatial information. On the other hand, if the segmentation is only performed via minimizing the weighted sum of class uncertainty, it will had that only the points on the strongest boundaries can get high class uncertainty. Moreover, this strategy cannot guarantee the global match between the model and the data (Xie and Tsui 2004). In order to solve the problem, in Xie (2004) is calculated the ϕ that maximize $g(y|\phi)$, the likelihood of parameters to the image. Meanwhile Xie (2004) pursues the purpose to make pixels with high class uncertainty accumulate mostly around the fuzzy boundaries. The loss function is defined as:

$$L = L_{entropy} - \lambda L_{likelihood} \tag{3.27}$$

where λ is a scale factor. The $L_{entropy}$ and $L_{likelihood}$ are the entropy and likelihood terms, respectively. The first term is defined as:

$$L_{entropy} = \sum_{i=1}^{N}\big(1 - G(x_i)\big)H(x_i) + \big(1 - H(x_i)\big)G(x_i) \qquad (3.28)$$
$$= \sum_{i=1}^{N}\big(1 - 2G(x_i)\big)H(x_i) + G(x_i)$$

where G is the normalized GVF value of pixels, H is the normalized entropy of pixels. Meanwhile, the likelihood term is defined as:

$$L_{likelihood} = \sum_{i=1}^{N} \log g(x_i) \qquad (3.29)$$

Then, only when both G and H are high or low, $L_{entropy}$ is low, thus the low value of the entropy term indicates high-class uncertainty is assigned to pixels with high GVF values or low uncertainty to pixels with low GVF values (Xie and Tsui 2004).

The maximum likelihood algorithm combined with a GVF field was proposed for to perform segmentation by Xie (2004). Gradient vector flow is a type of external force and GVF fields are calculated by applying generalized diffusion equations to both components of the gradient. This computation causes diffuse force vectors to exist far from the object, and crisp force vectors near the edges (Chen 2006).

Another supervised clustering algorithm is the supervised artificial neuronal-networks (ANNs) that are non-linear statistical data modeling tools (Wismuller, Vietze et al. 2000; Ma, Tavares et al. 2010). Radial basis function neuronal networks (RBFN) are an excellent prequel-type network, because of its good theoretical foundation, simple structure, fastness, and it has been always closely focused and widely applied. Neuronal network has naturally become the technology that can be combined with supervised clustering. RBFN has a good performance and the training of its hidden nodes is usually achieved with the help of the clustering, therefore RBFN and supervised clustering have the possibility of combining and improving with each other. For neuronal network model, the regression analysis was originally a concept in statistics and the floating of variables under test often can be attributed to other variables. While the regression method quantifies the relationship between the independent variables and non-independent variables, the method of variance analysis is suitable for dealing with problems having independent variables. In the initialization process of the RBFN algorithm, the clustering process is extended to the output space, and form a combined input-output space.

In the combined space by unsupervised clustering is obtained cluster centers, and the membership matrix are regarded as the initial value of the next phase. Through the initialization, the input data can be divided into c clusters, and for each cluster can be build the following linear regression model:

$$\hat{y} - z_i = a_i^T (x - v_i) \quad i = 1, 2, \ldots, c, \tag{3.30}$$

where a_i^T is the regression parameter of the i^{th} regression model; v_i and z_i are separately the locations of the cluster centers in the input space and output space. All the sample points and cluster centers belong to the same linear regression model and this combination represents the formation of the linear relationship. Through the value z_i, can also be obtained the error of the linear regression model. The objective function of the supervised clustering is to control the degree of impacts of the supervised item to the clustering process and adds an impact factor a. This impact factor of the supervised item has a great impact on the training result. When the impact factor is too large is easy to make the algorithm "bump" to be difficult to constrict, while it is too small, the supervised effect is not obvious. An initial value a_0 is defined manually and then according to a decreasing function in each iteration it is to reduce the value of a. The following function is responsible to adjust the value a:

$$a(t) = \frac{a_0}{1=100\frac{t}{T}}, \tag{3.31}$$

where *T* is the maximum number of times of iterations and *t* is the number of times of the current iteration. The next step of the RBFN algorithm is to calculate the regression function and then calculate a new clustering center. This process is repeated until the algorithm converges. Experiments in Zhu (2010) show that using this method train the RBF network and can obtain higher training accuracy.

Others supervised clustering algorithms frequently used, are the support vector machines and support vector machines algorithm, also known as maximum margin classifiers, because the algorithms try to find a hyperplane to maximize the margin between two classes and can simultaneously minimize the empirical classification error and maximize the geometric margin between classes (Ma, Tavares et al. 2010).

In medical images, the segmented regions often do not correspond well to true objects, because of the noise, large internal intensity variations in the structures to be identified and the overlapping of the distribution of intensity values corresponding to one structure

with those of another structure. To solve these problems, the known shape information on the desired objects is integrated into the segmentation method. Support vector machines (SVM) have been considered good candidates to solve the image segmentations problems because of its high generalization performance and sparse solution.

Zhao (2006) uses SVM density estimator to construct a prior knowledge model of the structure based on previously segmented training data. SVM is mainly developed to solve the classification problem, regression problem and density estimation problem. This algorithm is a new type of learning machines based on statistical learning theory (Zhao, Su et al. 2006). Furthermore, SVM is a powerful machine learning technique for binary classification supported by a strong theoretical foundation and excellent empirical success. The SVM's output function is expressed as:

$$f_{SVM}(x) = w^T \phi(x) + b \tag{3.32}$$

where superscript T represents transpose, and w and b are weight vector and bias, respectively. Input vector x is classified according to $y = \theta[f_{SVM}(x)]$, where $\theta[z]$ is 1 (one) for $z \geq 0$, and 0 (zero) otherwise. Geometrically, w indicates the normal vector of a hyperplane. All vectors $\phi(x)$ lying on one side of the hyperplane are classified as 1 (one) and all vectors on the other side as 0 (zero). Therefore, the objective of SVM is to find the parameters w and b for the optimal hyperplane to maximize the margin $\frac{1}{\|w\|}$ (distance between the hyperplane and the closest training instances termed support vector). Finally, the SVM's output function is obtained as:

$$f_{SVM}(x) = \sum_{i=1}^{M} \eta_i (2y_i - 1) K(x_i, x) + b \tag{3.33}$$

where η_i is the Lagrange multiplier for the first inequality constraint in equation and K is the kernel function.

Accordingly, for ANNs and SVMs the information extracted from the training set provides the features of structure in the form of weights or parameters that can be used for the later segmentation (Ma, Tavares et al. 2010).

The active shape models algorithms are nominated as supervised clustering algorithm and these algorithms use a shape model to generate new shapes (Bruijne, Ginneken et al. 2002). The active shape model (ASM) algorithm can also be called the landmark-based active shape model and it is based on the point distribution model (PDM). This

point distribution model is the basis for many other landmarks-based variants, which are used in the segmentation of structures. The aim of PDM is to build a model of the shape of the structure of interest that describes both typical shape and variability using examples of the training set. In the PDM, shapes are represented by a set of points or landmarks that are manually placed on each image and the labeling of the points is important. Furthermore, each labeled point represents a particular part of the object or its boundary (Duchesne, Pruessner et al. 2002). The process works by modeling how different labeled points tend to move together as the shape varies. If the labeling is incorrect, the method will fail to capture shape variability reliably. Moreover, if a point is not in the correct position on each shape, the model will be unable to precisely represent the position of that point. In the model is included terms that describes the noise caused by error in point location, and these points are manually identified during the training phase. So, landmarks for a given shape do not move independently and their positions are partially correlated (Bruijne, Ginneken et al. 2002; Duchesne, Pruessner et al. 2002). Bruijne (2002) proposes a modified ASM scheme in which optimal landmark positions are defined by maximum grey value profile correlation with adjacent slices rather than by correlation with profiles from the training data (Bruijne, Ginneken et al. 2002).

In the original ASM method, where in the first no information of neighboring slices is used, in both modified schemes no model information is used, then as a consequence segmentation errors are propagated through the dataset in the slice correlation schemes. Furthermore, to improve overall performance on the neighboring slice can be used a hybrid scheme combining both model information. The dependency on image slicing and orientation restricts the use of this method to segmentation tasks were the shape change between slices is not too large and the object is always imaged in approximately the same direction. However, many medical images are made using fixed scan protocols which satisfy these conditions (Bruijne, Ginneken et al. 2002).

Cootes (1998) developed the active appearance models (AAM) in part to improve the ASM method and make more use of the intensity information in the images, but also with the goal to match a full, photorealistic model directly on a new image instance by minimizing the difference between the image under interpretation and one synthesized by the model built. In this algorithm, the shape is incorporated into the AAM via a PDM and in the PDM the intensity under landmarks is sampled and used to generate a gray-level model. Then with this process is created two models, the gray-level and PDM, are concatenated, and a supermodel is created from principal components analysis (PCA) (Duchesne, Pruessner et al. 2002). PCA is applied to reduce the dimensionality of the data and with which shapes are described by their deviation from the mean shape (Bruijne, Ginneken et al. 2002). The resulting principal components can be considered as eigenmodes of appearance variation, embedding shape and intensity variability, and it is possible afterward to generate a new image instance each time the shape is deformed along those principal modes of variations. The synthesized image matched to

the original image and it is observed that shapes are thus modified to reduce a cost function, in the least-squares sense, between the synthesized and original images, rather than moving the shape points on an imaginary normal to the boundary. Since it matches known gray-level values within contiguous neighborhoods, the process becomes more robust. For performed AAM, in lieu of the PDM, it is proposed that a warp distribution model (WDM). This model is based on a statistical analysis of dense 3D deformation fields and it is able to effectively characterize shape. However, at the same time the WDM making use of all voxels within an image and thus avoid subsampling.

The appearance based (AB) segmentation is based on AAM algorithm and was introduced by Duchesne (2002). This algorithm is used to minimize the difference between a new image and one synthesized by the appearance model. Furthermore, the goal of the AB segmentation algorithm is to minimize the magnitude of the difference. The first step in this approach consists of building a linear relationship between variations in appearance parameters and gray-level synthesized images. The second step is to use an iterative algorithm to generate new estimates of the synthesized image that gradually approximated the new image. Through varying model parameters along each vector, the algorithm finds the closest match in the least-squares sense by minimizing the magnitude of the difference vector. Furthermore, by varying the parameters for each eigenvector new synthesized image instances can be calculated. Then, the closest matching point is found by finding the minimum of a second-order fit to the error matching function (Duchesne, Pruessner et al. 2002)

Unsupervised classification techniques are also called clustering algorithms and, with these techniques, the structure features are extracted from the classified points. Unsupervised classification family includes fuzzy C-means algorithms (FCM), iterative self-organizing data analysis technique algorithms (ISODATA) and unsupervised neuronal networks (Ma, Tavares et al. 2010).

Fuzzy C-means algorithms have been widely used in the image segmentation, and such a success is attributed to introduction of fuzziness for the belongingness of each image pixel, which makes the clustering methods able to retain more information from the original image than the crisp or hard segmentation methods. This clustering technique is achieved by iteratively minimizing a cost function that is dependent on the distance of pixels to the cluster centers in the feature domain. Furthermore, the spatial relationship of neighboring pixels is an important characteristic that can be of great aid in image segmentation (Li and Shen 2010). The fuzzy C-means clustering functions well on images with low levels of noise, but has problems when used in segmentation of noise corrupted images. In one hand, the FCM does not incorporate the information concerning the spatial context, which makes it sensitive to noise and other imaging artifacts. On the other hand, the cluster assignment is based exclusively on the distribution of the pixel intensity that makes it sensitive to intensity variations due to illumination conditions or object shape. Many algorithms have been presented in order to improve the robustness of conventional FCM, and these methods can be classified

into two main groups: imposing spatial constraints to clustering algorithms and introducing other features or dissimilarity index that is insensitive to intensity variations in the objective function of FCM (Beevi 2010).

The standard FCM algorithm assigns pixels to each category by using fuzzy memberships, and it is an iterative optimization that minimizes the cost function defined as:

$$J_m(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} d_{ij}^{2}, \tag{3.34}$$

$$u_{ij} \in [0,1]\,; \quad 1 \le j \le n\,; \quad 1 \le i \le c, \tag{3.35}$$

where $m \in (1,\infty)$ and m factor controls the fuzziness of the resulting partition ($m = 2$ in Li (2010) study), u_{ij} represents the membership of pixel x_j in the i^{th} clustering, $d_{ij} = \|x_j - v_i\|$ represents the distance between the pixel x_j and the cluster center v_i. The gray-level value is the most commonly used feature in the image clustering. Thus, the FCM cost function is minimized when high membership values are assigned to pixels whose intensities are close to the centroid. Moreover, the probability that a pixel belongs to a specific cluster represents membership function and this value is dependent solely on the distance between the pixel and each individual cluster center domain. Starting with an initial guess for each cluster center, the FCM converges to a solution for v_i representing the local minimum or a saddle point of the cost function and convergence can be detected by comparing the changes in the membership function or the cluster center at two successive iteration steps. Then, after the convergence, defuzzification is applied to assign each pixel to a specific cluster for which the membership is maximal (Li and Shen 2010).

Recently, the performance of image segmentation with the FCM algorithm was improved by the incorporation of spatial information into the standard FCM method (Beevi 2010).

The unsupervised classification also includes the k-means algorithm, also called iterative self-organizing data analysis technique algorithm. The k-means algorithm clusters data by iteratively computing a mean intensity for each class and segmenting the image by classifying each pixel in the class with the closest mean (Pham, Xu et al. 2000; Ma, Tavares et al. 2010). Furthermore, the k-means algorithms are known to be fast solutions for solving such problem. However, they are sensitive to the choice of starting points and can only be applied to small datasets. To solve these disadvantages, Bagirov (2011) proposed the global k-means algorithms and modified k-means

algorithm and the results of numerical experiments in (Bagirov, Ugon et al. 2011) show that these algorithms allow one to find global or near global minimizer of the clusters of the cluster function. However, these algorithms are memory demanding as they require the storage of the affinity matrix. Alternatively, this matrix can be computed at catch iteration, but this extends the computational time significantly. The k-means algorithm is applied from starting points lying in the different parts of the dataset to minimize the auxiliary cluster function, and the best solution is selected as a starting point for the next cluster center. The modified global k-means algorithm is a new version of the k-means algorithm that reduces the amount of computational effort by removing data points which are close to cluster centers found in the previous iteration and using the triangle inequality for distances to avoid unnecessary computations. The modified global k-means algorithm is an incremental clustering algorithm and to solve k-partition problem it starts with the computation of one cluster, that is with the computation of the centroid of the dataset, and attempts to optimally add one new cluster center at each iteration. Thus, this algorithm computes cluster incrementally, using the k-1 cluster centers from the previous iteration to solve the k-partition problem. An important step in this algorithm is the computation of a starting point for the k-th cluster center, and this starting point is computing by minimizing the so-called auxiliary cluster function. Unlike the modified global k-means algorithm the proposed algorithm does not rely on the affinity matrix to compute the starting point. Furthermore, the auxiliary cluster function is nonconvex and it may have many local solutions. One can expect that the global minimizer of this function is a good candidate as starting point for the k-th cluster center (Bagirov, Ugon et al. 2011).

As already referred, the unsupervised methods explore the intrinsic data structure to segment the image into regions with different statistics. However, these methods often fail to achieve the desired result, especially if the desired segmentation includes regions with very different characteristics. On the other hand, supervised image segmentation methods first learn a classifier from a labeled training set. Although these methods are likely to perform better, marking the training set is very time consuming. Semi-supervised image segmentation is the last type of clustering technique, and these methods circumvent these problems by inferring the segmentation from partially labeled images. Thus, the key difference from supervised learning is that semi-supervised methods utilize the data structure in both labeled and unlabeled data points. Furthermore, the main advantage of semi-supervised image segmentation methods is that they take advantage of the users' markings to direct the segmentation, while minimizing the need for user labeling (Figueiredo 2007; Xiao-min, Xiao et al. 2009; Paiva and Tasdizen 2010). Recent developments of semi-supervised approaches have focused on graph-based methods, because the graph-based representation naturally copes with nonlinear data manifolds and with this formulation, data points are represented by nodes in a graph, and the edge weights are given by some measure of distance or affinity between data points. However, results from graph-based semi-supervised methods are severely limited by the number of data points, because label

propagation in the graphs requires first the computation of connectivity matrix and the computational complexity grows exponentially with the number of points. Therefore, this severely limits the application of semi-supervised learning methods for image segmentation (Paiva and Tasdizen 2010).

Most of the methods for semi-supervised clustering work by incorporating the desired relations into classical algorithms such as the expectation-maximization algorithm for mixture-based clustering or the K-means algorithm. The desired relations may be imposed in a hard way or used to build priors under which probabilistic clustering is performed. Figueiredo (2007) described a new approach to semi-supervised mixture-based clustering for which he derives a simple, fully deterministic generalized EM (GEM) algorithm. The keystone of this approach is the formulation of the problem, where the labels are only indirectly observer and the linearity of the resulting complete log-likelihood, with respect to the missing group labels, will allow deriving a simple GEM algorithm.

To conclude, the semi-supervised clustering is the recent type of clustering that takes advantage of the users' markings to direct the segmentation, while minimizing the need for user labeling, and has been applied to problems of symptoms classification in medical images and promising results have been obtained (Figueiredo 2007; Xiao-min, Xiao et al. 2009; Li and Shen 2010; Paiva and Tasdizen 2010).

3.2.3 Algorithms based on Deformable Models

When compared with the previous types described, algorithms based on deformable models are more flexible, and can be used for complex segmentation problems (Ma, Tavares et al. 2010).

The designation "deformable models" stems primarily from the use of elasticity theory at the physical level, generally within a Lagrangian dynamics setting. The physical interpretation views deformable models as elastic bodies which respond naturally to applied forces and constraints. Typically, deformation energy functions defined in terms of the geometric degrees of freedom are associated with the deformable model. The energy grows monotonically as the model deforms away from a specified natural or "rest shape" and often includes terms that constrain the smoothness or symmetry of the model. In the Lagrangian setting, the deformation energy gives rise to elastic forces internal to the model (McInerney and Terzopoulos 1996).

The widely recognized potency of deformable models stems from their ability to segment, match, and track images of anatomic structures by exploiting (bottom-up)

constraints derived from the image data together with (top-down) a *priori* knowledge about the location, size and shape of these structures. Deformable models are capable of accommodating the often significant variability of biological structures over time and across different individuals. Furthermore, deformable models support highly intuitive interaction mechanisms that allow medical scientists and practitioners to bring their expertise to bear on the model-based image interpretation task when necessary (McInerney and Terzopoulos 1996; Pham, Xu et al. 2000; Ma, Tavares et al. 2010). The main advantages of deformable models are their ability to directly generate closed parametric curves or surfaces from images and their incorporation of a smoothness constraint that provides robustness to noise and spurious edges. A disadvantage is that they usually require manual interaction to place an initial model and choose appropriate parameters (Pham, Xu et al. 2000).

According to the way that is used for tracking the moving contour, deformable models can be further classified into parametric models and geometric models (Ma, Tavares et al. 2010). Parametric deformable models represent curves and surfaces explicitly in their parametric forms during deformation. This representation allows direct interaction with the model and can lead to a compact representation for fast real-time implementation. Adaptation of the model topology, however, such as splitting or merging parts during the deformation, can be difficult using parametric models. Geometric deformable models, on the other hand, can handle topological changes naturally. These models, based on theory of curve evolution and the level set method, represent curves and surfaces implicitly as a level set of a higher-dimensional scalar function. Their parameterizations are computed only after complete deformation, thereby allowing topological adaptively to be easily accommodated. Despite this fundamental difference, the underlying principles of both methods are very similar (McInerney and Terzopoulos 1996).

The main feature of parametric deformable models is the tracking of the evolution through sampled contour points. Explicit tracking has the advantage of high computational efficiency and allows for real-time applications. The moving equation for the contour can be derived through energy functions or defined directly through dynamic forces. There are two different types of formulations for parametric deformable models: an energy minimizing formulation and a dynamic force formulation. Although, these two formulations lead to similar results, the first formulation has the advantage that its solution satisfies a minimum principle, whereas the second formulation has the flexibility of allowing the use of more general types of external forces. The basic premise of the energy minimizing formulation of deformable contours is to find a parameterized curve that minimizes the weighted sum of internal energy and potential energy. The internal energy specifies the tension or the smoothness of the contour. On the other hand, the potential energy is defined over the image domain and typically possesses local minima at the image intensity edges occurring at object boundaries. Minimizing the total energy yields internal forces and potential forces. Moreover,

internal forces hold the curve together (elasticity forces) and keep it from bending too much (bending forces), while external forces attract the curve toward the desired object boundaries. In order to find the object boundary, parametric curves are initialized within the image domain, and are forced to move toward the potential energy minima under the influence of both these forces (McInerney and Terzopoulos 1996; Xu and Prince 1998).

A deformable contour is a curve, which moves through the spatial domain of an image to minimize the following energy functional:

$$\mathcal{E}(X) = \mathcal{S}(X) + \mathcal{P}(X) \tag{3.36}$$

where the first term is the internal energy functional and is defined to be:

$$\mathcal{S}(X) = \frac{1}{2}\int_0^1 \alpha(s)\left|\frac{\partial X}{\partial s}\right|^2 + \beta(s)\left|\frac{\partial^2 X}{\partial s^2}\right|^2 ds \tag{3.37}$$

and the first-order derivate discourages stretching and makes the model behave like an elastic string, while the second-order derivate discourages bending and makes the model behave like a rigid rod. Parameters $\alpha(s)$ e $\beta(s)$ can be used to control the strength of the model's tension and rigidity, respectively.

The second term in Eq. 3.36 is the potential energy functional and is computed by integrating a potential energy function along the contour $X(s)$:

$$\mathcal{P}(X) = \int_0^1 P\big(X(s)\big)ds \tag{3.38}$$

where $P(x, y)$ is the potential energy function. This function is derived from the image data and takes smaller values at object boundaries as well as other features of interest, and when is given a gray-level image $I(x, y)$, the typical potential energy function designed to lead a deformable contour toward step edges is:

$$P(x, y) = -w_e|\nabla[G_\sigma(x, y) * I(x, y)]|^2 \tag{3.39}$$

where w_e is a weighting parameter, $G_\sigma(x,y)$ is a 2D Gaussian function with standard deviation σ, ∇ is the gradient operator, and $*$ is the 2D image convolution operator. If the desired image features are lines, then the appropriate potential energy function can be defined as:

$$P(x,y) = w_1[G_\sigma(x,y) * I(x,y)], \tag{3.40}$$

and w_1 is also a weighting parameter. When w_1 is positive, it is used to find black lines on a white background, while if w_1 is negative, it is used to find white lines on a black background. For both edge and line potential energies, increasing σ can broaden its attraction range. However, larger σ can also cause a shift in the boundary location, resulting in a less accurate result (McInerney and Terzopoulos 1996; Xu and Prince 1998).

To gain some insight about the physical behavior of deformable contours, we can view as a force balance equation:

$$F_{int}(X) + F_{pot}(X) = 0, \tag{3.41}$$

where the internal force is given by:

$$F_{int}(X) = \frac{\partial}{\partial s}\left(\alpha \frac{\partial X}{\partial s}\right) - \frac{\partial^2}{\partial s^2}\left(\beta \frac{\partial^2 X}{\partial s^2}\right), \tag{3.42}$$

and the potential force is given by:

$$F_{pot}(X) = -\nabla P(X). \tag{3.43}$$

The potential force $F_{pot}(X)$ pulls the contour toward the desired object boundaries, while the internal force $F_{int}(X)$ discourages stretching and bending (McInerney, 1996; Xu, 1998).

The deformable contour is made dynamic by treating $X(s)$ as a function of time t:

$$\gamma \frac{\partial X}{\partial t} = \frac{\partial}{\partial s}\left(\alpha \frac{\partial X}{\partial s}\right) - \frac{\partial^2}{\partial s^2}\left(\beta \frac{\partial^2 X}{\partial s^2}\right) - \nabla P(X) \tag{3.44}$$

where the coefficient γ is introduced to make the units on the left side consistent with the right side. Thus, the minimization is solved by placing an initial contour on the image domain and allowing it to deform according to Eq. (3.44) (McInerney and Terzopoulos 1996; Xu and Prince 1998).

As shown before, the deformable model can be modeled as a static problem, and an artificial variable t was introduced to minimize the energy. However, it is sometimes more convenient to formulate the deformable model directly from a dynamic problem using a force formulation. Such a formulation permits the use of more general types of external forces that are not potential forces, for example, forces that cannot be written as the negative gradient of potential energy functions. The dynamics of a contour $X(s,t)$ must satisfy:

$$\mu \frac{\partial^2 X}{\partial t^2} = F_{damp}(X) + F_{int}(X) + F_{ext}(X), \tag{3.45}$$

where μ is a coefficient that has a mass unit and F_{damp} is the damping force (or viscous force) defined as $\gamma \frac{\partial X}{\partial t}$, with γ being the damping coefficient. The mass coefficient μ in front of the inertial term for the image segmentation is often set to zero, since the inertial term may cause the contour to pass over the weak edges. Thus, the dynamics of the deformable contour without inertial term becomes:

$$\gamma \frac{\partial X}{\partial t} = F_{int}(X) + F_{ext}(X), \tag{3.46}$$

and the external forces can be either potential forces or nonpotential force. However, nonpotential forces cannot be derived from the variational energy, but they are often expressed as the superposition of several different forces. This superposition formulation allows the external forces to be broken down into more manageable terms. For example, one might define the external forces to be composed of both Gaussian potential forces and pressure forces. For using the Gaussian potential force, σ must be selected to have a small value in order to the deformable model to follow the boundary accurately. As a result, the Gaussian potential force can only attract the model toward the boundary when it is initialized nearby. On the other hand, the pressure force can

either inflate or deflate the model, and then deformable models that use pressure forces are also known as balloons (McInerney and Terzopoulos 1996; Xu and Prince 1998).

The snake method was the first deformable model applied to the medical image segmentation, and the development of parametric deformable models has a tight relationship with the snake method. The original snake method used the tension and rigidity of the contour, the internal energy, the gradient magnitude as the external energy. Furthermore, the snake method is sensitive to the initial conditions. The moving contour may stop at places with local functional minimum or places where the gradient magnitude is too small so that the external forces tend to be zero. In the non-interactive applications, the snakes must be initialized close to the structure of interest to guarantee good performance. Consequently, in order to get a correct segmentation the initial contour must have the same topology as the desired object and must be placed near the object boundary (McInerney and Terzopoulos 1996; Ma, Tavares et al. 2010). Therefore, there are two key difficulties with parametric active contour algorithms. First, the initial contour must, in general, be close to the true boundary, or else it will likely converge to the wrong result. The basic idea to address this problem is to increase the capture range of the external force fields and to guide the contour toward the desired boundary. The second problem is that active contours have difficulties processing into boundary concavities.

In order to eliminate the dependence on the initial position Xu (1998) analyzed the reason why snake methods have poor convergence to boundaries with large curvatures and replaced the gradient field with the gradient vector field (GVF), which has a larger capture region and slowly changes away from the boundaries. These changes decreased the dependence on initial positions, but the field can attract the moving contour to right position (Ma, Tavares et al. 2010).

The gradient vector flow fields are dense vector fields derived from images by minimizing certain energy functional in a variational framework. The minimization is achieved by solving a pair of decoupled linear partial differential equations that diffuses the gradient vectors of a gray-level or binary edge map computed from the image. The active contour that uses the GVF field is called GVF snake. The GVF snake is distinguished from nearly all previous snake formulations in that its external forces cannot be written as the negative gradient of a potential function. Because of this, it cannot be formulated using the standard energy minimization framework; instead, it is specified directly from a force balance condition. Particular advantages of the GVF snake over a traditional snake are its insensitivity to initialization and its ability to move into boundary concavities. Furthermore, GVF snake does not need prior knowledge about whether to shrink or expand toward the boundary. The GVF snake also has a large capture range, which means that, barring interference from other objects, it can be initialized far away from the boundary. This increase capture range is achieved through a diffusion process that does not blur the edges themselves, so multi-resolution methods are not needed (Xu and Prince 1998).

Statistical techniques can be integrated in parametric deformable models and they are also popular. Typical examples include ASM and AAM and to extract the mean shape and to define proper ranges of the parameters are used training samples. ASM uses the edge information to move the shape points to better positions after finding an approximate position. On the other hand, AAM uses the mean texture of each shape point to find a better position. The searching procedure is like the snake method, but the movements of shape points are constrained by ranges of shape parameters which guarantee the similarity between the segmentation result and the training samples. This characteristic is very useful when the shape or topology of structures can hardly be identified from their appearances in the images (Ma, Tavares et al. 2010).

Parametric deformable models are widely used in structure segmentation and 3D reconstructions; however, the computational complexity such as parameterisation of the contours, handling of topological changes, and re-distribution of the contour points considerably restricts their applications.

In parametric deformable models an explicit parametric representation of the curve is used, which can lead to fast real-time implementation. However, it is difficult for parametric deformable models to adapt the model topology during deformation. On the other hand, geometric deformable models are designed to handle topological changes naturally. These models are based on the theory of curve evolution, and the level set method, where the evolving curves or surfaces are implicitly represented as a level set of higher-dimensional scalar function, a level set function. Furthermore, geometric deformable models are also called level set methods (Wang, Zhu et al. 2007). The main idea of the level set method is to implicitly embed the moving contour into a higher-dimensional level set function and view the contour as its zero level set. Then, instead of tracking the discrete contour points, one can track the zero level set of the level set function. The advantage of doing so is that the topological changes can be easily handled and the geometric properties of the contour can be implicitly calculated. Therefore, the complexity of geometric deformable models, speed functions should be defined properly to drive the contour to the right position (Ma, Tavares et al. 2010). Thus, the level set method views a moving curve as the zero level set of a higher-dimensional function $\phi(x,t)$. The level set function satisfies:

$$\begin{cases} \phi(x,t) < 0 \ \ in \ \ \Omega(t), \\ \phi(x,t) = 0 \ \ in \ \ C(t), \\ \phi(x,t) > 0 \ \ in \ \ R^n \setminus \bar{\Omega}(t), \end{cases} \tag{3.47}$$

where the artificial time t denotes the evolution process, $C(t)$ is the moving curve and $\Omega(t)$ represents the region that $C(t)$ encloses. The evolution equation of the moving curve with speed F in its normal direction is given by:

$$\phi_t = F(x)|\nabla\phi|. \tag{3.48}$$

In Eq. 3.48, the surface $\phi = 0$ corresponding to the propagating hypersurface may change topology, as well as form sharp corners.

The motion by mean curvature is a particular case of the level curve of ϕ passing through x and this particular case happens when $F = div(\frac{\nabla\phi(x)}{|\nabla\phi(x)|})$. The equation 3.48 becomes:

$$\frac{\partial\phi}{\partial t} = |\nabla\phi|.div\left(\frac{\nabla\phi}{|\nabla\phi|}\right), \tag{3.49}$$

with $\phi(0,x) = \phi_0(x)$ and $t \in (0,\infty)$.

Geometric deformable model based on mean curvature motion is given as:

$$\frac{\partial\phi}{\partial t} = g(|\nabla f|).|\nabla\phi|.\left[div\left(\frac{\nabla\phi}{|\nabla\phi|}\right) + v\right], \tag{3.50}$$

and, in this equation, $v \geq 0$ is a constraint on the area inside the curve, increasing the propagation speed, where $g(|\nabla f|)$ is an edge-sensitive speed function defined as:

$$F(x) = \frac{1}{1+|\nabla\{G_\sigma * f(x)\}|^p}, \tag{3.51}$$

where $p \geq 1$. With this argument is clear that, if the image gradient $|\nabla\{G_\sigma * f(x)\}|$ approaches the local maximum at the object boundaries, the curve gradually attains zero speed. The ideal condition is $|\nabla\{G_\sigma * f(x)\}| \to \infty$ and $F = 0$ at the boundaries, the evolving curve eventually stops and the final zero level set $\psi(x,\infty) = 0$ corresponds to the segmentation result.

Malladi's algorithm also uses the gradient information to define the speed function and add the curvature influence to keep the contour smooth. Malladi's speed function model is intuitive because when the contour moves to the structure boundary, the increase of the gradient magnitude decreases the speed value so; the evolution of the contour slows down. Furthermore, the evolution can be stopped after a time to gain the position of the

structure boundary, and the stopping criterion should be selected carefully to make sure the contour stops at the right position. However, the contour may leak or shrink to disappearance after a long evolution in the images with noise or blurred (Ma, Tavares et al. 2010).

The geodesic snake model is another well-known deformable model (Wang, Zhu et al. 2007). Unlike Malladi's model, the geodesic active contour (GAC) algorithm modeled the segmentation as an optimization problem of finding the minimal distance curve in the image. This model also uses the image gradient to stop the curve and its level set formulation is as follows:

$$\frac{\partial \phi}{\partial t} = |\nabla \phi|.\, div\left(g(|\nabla f|)\frac{\nabla \phi}{|\nabla \phi|}\right) + vg(|\nabla f|)|\nabla \phi|. \tag{3.52}$$

The moving equation of GAC is derived from the above function and the procedure of finding the optimal solution corresponds to the searching of the structure boundary. Moreover, unlike in Malladi's models, the equilibrium state of the moving contour guarantees that a long computation time will not lead to leakage (Wang, Zhu et al. 2007; Ma, Tavares et al. 2010). With the GAC algorithm can be seen a tight relationship between the parametric model and the geometric model. The contours of interest can be post-processed by using statistical techniques such as maximizing a posterior (MAP) and principle component analysis (PCA). Furthermore, to eliminate dependence on the initial conditions can be combined the gradient vector flow with the GAC algorithm and with this combination the results are better than the original GAC algorithm (Ma, Tavares et al. 2010).

The classical snake models rely on the edge function $g(|\nabla f|)$, i.e., depend on the image gradient $|\nabla f|$ to stop the curve evolution and these models can detect only objects with edges defined by a gradient. Therefore, in practice, discrete gradients are bounded, and so stopping function is never zero on the edges, and the curve may pass through the boundary, even for the geodesic snake model mentioned above. In case of the image to be very noisy, the isotropic smoothing Gaussian has to be strong. To solve these problems, Chan and Vese proposed a different deformable model that is not based on the gradient of the image for the stopping process. Chan and Vese's model is a simplified version of the Mumford and Shah energy model and the evolvement of the curve is based on the general Mumford-Shah formulation (Eq. 3.53) of the image segmentation, by minimizing the functional:

$$F^{MS}(f, C) = \mu Length(C) + \lambda \int_{\Omega} |f - f_0|^2 dxdy + \int_{\Omega/C} |\nabla f|^2 dxdy \tag{3.53}$$

In Eq. 2.52, $\vec{C}$ denotes the smooth, closed segmenting curve, μ and λ are positive parameters, $f_0: \overline{\Omega} \longrightarrow R$ is a given image, Ω is the image domain. The solution image f is formed by smooth regions R_i and sharp boundaries, denoted here by C (Tsai, 2001; Wang, 2007).

The restriction F^{MS} can be reduced to piecewise constant functions and then can be finding a partition of Ω such that f in Ω_i equals a constant. Based on the above equation, Chan and Vese proposed the following minimization problem for two-phase segmentation:

$$\min_{C,c_0,c_b}\left\{\mu\int_{\Omega}\delta(\phi)|\nabla\phi|dxdy + v\int_{\Omega}H(\phi)dxdy + \lambda_0\int_{inside(C)}|f-c_0|^2H(\phi)dxdy + \lambda_b\int_{outside(C)}|f-c_b|^2\big(1-H(\phi)\big)dxdy\right\} \tag{3.54}$$

where ϕ is the level set function and $H(\phi)$ is the Heaviside function:

$$H(\phi)=\begin{cases}1, & \phi \geq 0,\\ 0, & \phi < 0.\end{cases} \tag{3.55}$$

Usually, the deformable models implemented by means of the level set method suffer from a slower speed of convergence than parametric deformable models due to their computational complexity. However, they can automatically handle topology changes and allow for multiple simultaneous boundary estimations. Specifically, algorithms based on geometric deformable models aim to eliminate noise influence, prevent leakage, enhance accuracy and efficiency, and make the algorithms more automatic and less dependent on the initial conditions (Tsai, 2001; Wang, 2007).

In conclusion, deformable models are promising for the segmentation of the medical images, because these models can easily incorporate statistical information and other techniques, while using curve evolution to find the optimal boundaries can provide a contour with regular geometric properties.

3.3 Summary

Image segmentation is a basic problem in Computational Vision and the key to the procedure from processing to analyzing. Extracting object contour of images, obtaining specific edge information that can assist the image interpretation. Furthermore, segmentation is inserted in the analysis stage and can be divided into three classes based on their essential techniques: thresholding, clustering and deformable models.

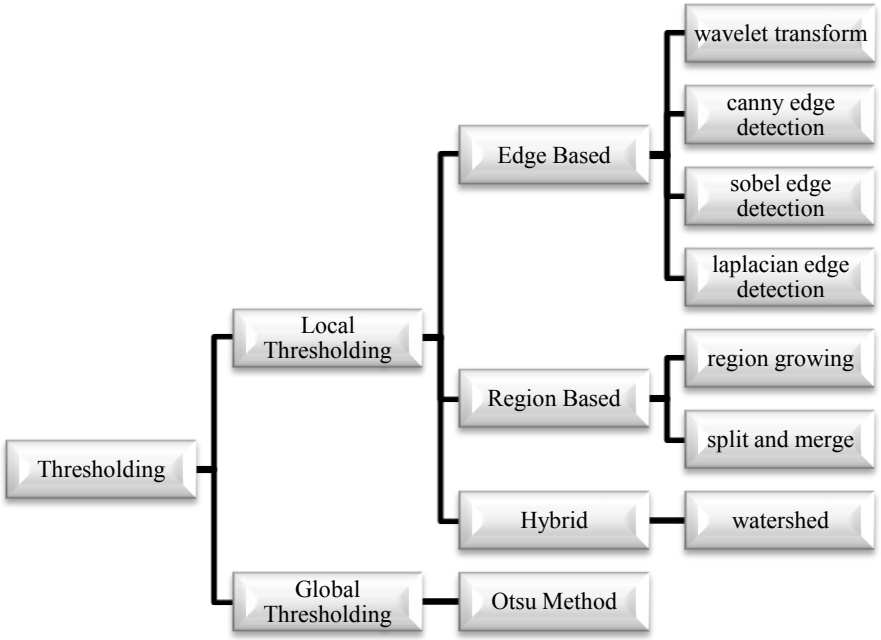


Figure 3.6: Different thresholding-base segmentation algorithms.

There are several thresholding techniques and some of them are based on the image histogram, global Thresholding (e.g. Otsu method); others are based in local properties or local gradient, local Thresholding. Moreover, local thresholding algorithms can be further classified as edge-based ones, region-based ones and hybrid ones, Figure 3.19.

The clustering algorithms have the same function as classifier methods and classifier methods are pattern recognition techniques that seek to partition a feature space derived from the image using data with known labels (Pham, 2000). In clustering algorithms, the partitioning is based on pairwise relations between individual patterns, rather centralized relations between samples and a few cluster representatives (like centroids, for example) (Matta and Dugelay 2009). Furthermore, the clustering techniques can be divided in three main parts: supervised algorithms, unsupervised algorithms and semi-supervised algorithms, Figure 3.20 (Sutton, Bezdek et al. 2000; Ma, Tavares et al. 2010). The supervised techniques are frequently used and include k-nearest neighbor (kNN), maximum likelihood (ML) algorithms, supervised artificial neuronal networks (ANN), support vector machines (SVM), active shape models (ASM) and active

appearance models (AAM). On the other hand, unsupervised classification techniques are also called clustering algorithms, and with these techniques, the structure features are extracted from the classified points. Unsupervised classification includes fuzzy C-means algorithms (FCM), iterative self-organizing data analysis technique algorithms (ISODATA) and unsupervised neuronal networks (Ma, Tavares et al. 2010). Finally, semi-supervised image segmentation is the last type of clustering technique and these methods circumvent these problems by inferring the segmentation from partially labeled images. Thus, the key difference from supervised learning is that semi-supervised methods utilize the data structure in both labeled and unlabeled data points. Furthermore, the main advantage of semi-supervised image segmentation methods is that they take advantage of the users' markings to direct the segmentation, while minimizing the need for user labeling (Figueiredo 2007; Xiao-min, Xiao et al. 2009; Paiva and Tasdizen 2010).

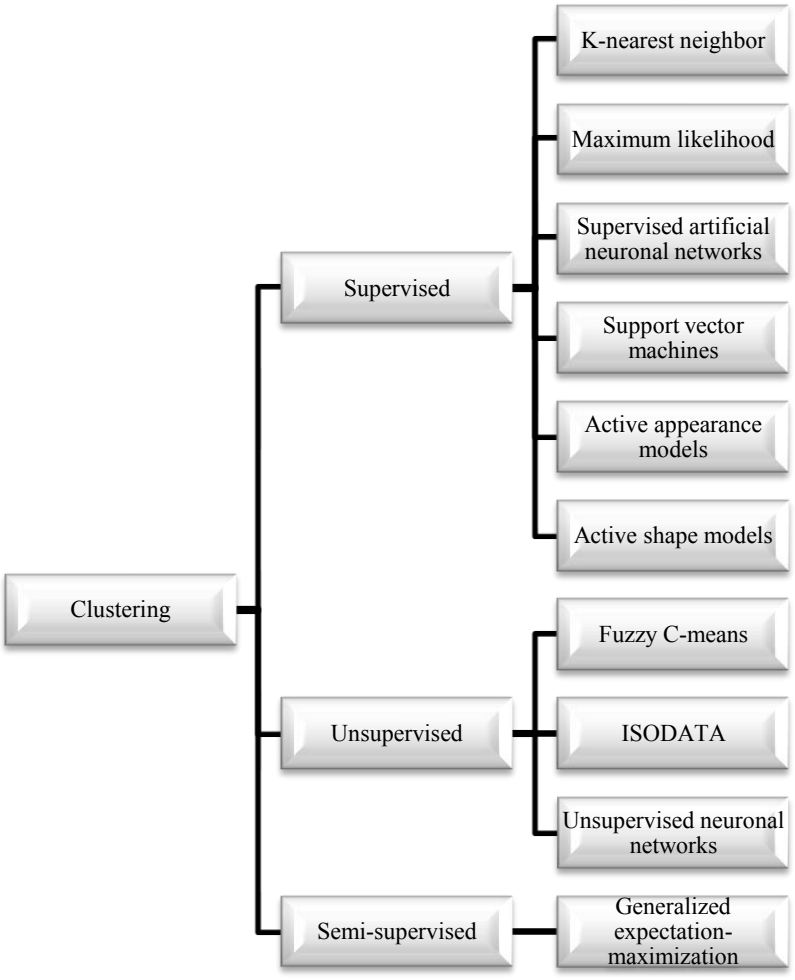


Figure 3.7: Image segmentaion algorithms based on clustering methods.

Algorithms based on deformable models are more flexible when compared with the above two types, and they can be used for complex segmentations and according to the

way that is used for tracking the moving contour, deformable models can be further classified into parametric models and geometric models, Figure 3.21.

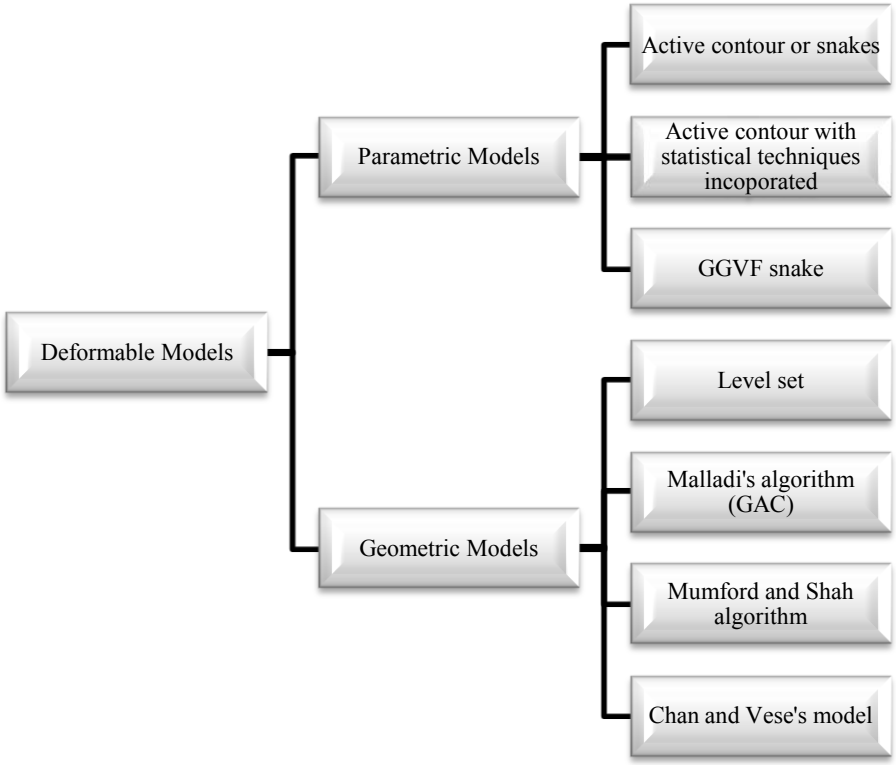


Figure 3.8: Segmentaion algorithms based on deformable models.

Chapter IV - Segmentation Algorithms for Human Ear Images

- Introduction;
- Segmentation of the Tympanic Membrane;
- Segmentation of the Middle Ear;
- Segmentation of the Inner Ear;
- Summary.

4.1 Introduction

The Human auditory system consists of the organs of hearing and balance and is usually divided into three parts: external ear, middle ear and inner ear. The ear as a sensory organ is far more complex than other sensory organs, and is used to warning and to communicate. The sensory cells are located in the cochlea, but the cochlea not only serves to convert sound into a code of neural impulses in the auditory nerve, but it also performs the first analysis of sounds that prepare sounds for further analysis in the auditory nervous system (Seeley, Stephens et al. 2004; Moller 2006).

A number of approaches have been presented for building 3D models of the ear for education (Jun, Song et al. 2005; Wang, Zhu et al. 2007), biomechanical studies (Decraemer, Dirckx et al. 2003; Sim and Puria 2008) and pre-operative assessment (Hussong, Rau et al. 2009; Rau, Hussong et al. 2010). In order to perform these approaches can be used different types of medical images that are obtained from several imaging techniques, such as, Computerized Tomography (CT-standard, Micro-CT, Spiral-CT) (Christensen, He et al. 2003; Xianfen, Siping et al. 2005; Poznyakovskiy, Zahnert et al. 2008), Magnetic Resonance (MR-standard, Micro-MR) (Lane, Witte et al. 2005; Liu, Gao et al. 2007; Shi, Wang et al. 2010) and Histological processing (Liu, Gao et al. 2007). To study the anatomical features of the ear have been used cat, guinea pig, chinchilla and human ears (Liu, Gao et al. 2007; Sim and Puria 2008).

Since the late 1980's, several researches have combined the technology of image processing, medical imaging and histological process to model the structures of the cochlea. So far, modeling of the middle ear has proven to be quite accurate and vivid when compared to the actual morphology of the middle ear (Liu, Gao et al. 2007).

The image processing technology is essential for closer the geometric models to anatomical structures of the ear. Thereby, the segmentation of the ear structures is crucial to creating computational models and, when these models are used, it is possible to perform simulations of the ear structures. From this simulation becomes easier understanding the connections between the ear structures and their functions. For example, the using of geometric models of the ear structures makes easier the understanding and the optimization of the prosthetic implants. On the other hand, also are improved the methodologies for planning the surgical procedures (Tuck-Lee, Pinsky et al. 2008). The biomechanical modeling of the ear also presents a key role in diagnosis and treatment of middle and inner ear diseases, because these two processes are hampered by the small size of the structures and by the hidden locations of the middle and inner ear structures in the temporal bone (Seemann et al. 1999). In addition, through the modeling of the inner ear can be identified anatomic abnormalities of the bony labyrinth, 15% of patients with hearing loss have gross configurationally anomalies. Therefore, there is a possibility of to create templates that standardize the abnormal configurations (Melhem, Shakir et al. 1998). Cochlear implant systems present as a

direct application of the geometric modeling of the ear, especially today, because they are used in clinical interventions in patients with deafness or severe hearing loss. The position of the implanted electrodes have been present as one the most important variables in the speech recognition. Through the geometric modeling is possible to create automation in the electrode insertion process. This process is an important step towards a traumatic cochlear implant surgery. Up to now, only manual insertion tools or insertion aids exist and these tools provide the possibility to manually insert the electrode with one given, fixed insertion technique which is not adjustable to the patient's need (Hussong, Rau et al. 2009; Rau, Hussong et al. 2010). In addition, the post processing, as the virtual endoscopy technique has been highly developed as a form of simulation of the middle ear surgery (Liu, Gao et al. 2007).

Following, the use of computational algorithms to segment ear structures in images is presented. The algorithms to be presented have into account the image type to be analyzed, the characteristics of the shape of the structures, the texture characteristics and the intensities range of the region of interest.

4.2 Segmentation of the Tympanic Membrane

As already mentioned, it is known that the ear is divided in three parts: external ear, middle ear and inner ear. The structure that divides the external ear from the middle ear is the tympanic membrane and it is externally observed from video-otoscopy and otoscopy images (Xie, Mirmehdi et al. 2005; Comunello, Wangenheim et al. 2009). Through analysis of these types of images can be diagnosed several pathologies, such as perforations, otitis media and cholesteatoma (infection within the middle ear cleft) (Xie, Mirmehdi et al. 2005).

One of the algorithms that can be used to perform the segmentation of this structure is the generalized version of the well-known gradient vector flow snake (GGVF) (Xie, Mirmehdi et al. 2005). Xie, Mirmehdi et al. (2005) used this algorithm with the aim to delineate the tympanic membrane boundaries and to detect color abnormalities in the tympanic membrane. This geometric generalized GVF snake is useful to delineate boundaries with small gaps and tympanic membrane boundaries present this feature. The GGVF snake presents other advantages that demonstrate its efficiency to segment the tympanic membrane boundaries, such as insensitivity to initialization, ability to move into boundary concavities and a large capture range that barring interference from other objects. This increased capture range is achieved through a diffusion process that does not blur the edges themselves, multi-resolution methods are not needed (Xu and Prince 1998).

The tympanic membrane also can be segmented by the Mumford-Shah's semi-automatic algorithm (Comunello, Wangenheim et al. 2009), Figure 4.1.

Figure 4.1: An example of the Mumford-Shah segmentation results (from (Comunello, Wangenheim et al. 2009))

The Mumford-Shah algorithm was used by Comunello et al. (2009) to analyze quantitatively the tympanic membrane and its pathologies. This algorithm presents effective in image segmentation of the tympanic membrane, because it has high robustness in the presence of noise and in the choice of place to start the segmentation (Tsai and Willsky 2001). In addition, this algorithm guarantees that no segment leakage between structures occurs, and it also allows knowledge the quantitative information about tympanic membrane perforations, so it is indicated to the clinic diagnosis (Comunello, Wangenheim et al. 2009).

4.3 Segmentation of the Middle Ear

The middle ear is characterized by the bone structures and, theoretically, would be possible to segment these structures using a thresholding algorithm by selecting a value between the air intensity value (low value), and the bone intensity value (high value). However, the partial volume effects suppress the bone structures that are around the inner ear and, thus the middle and inner ear bone structures present a similar intensity distribution.

Thresholding algorithm was used by Lee et al. (2010) to segment the middle ear in Micro-CT images that present a higher anatomical resolution. In addition, such images

allow a detailed observation of the middle ear structures that have a reduced size (Lee, Chan et al. 2010). The thresholding algorithm is also used by Rodt et al. (2002), but, to reduce the noise, morphologic operations (erosion and dilation) were applied on the input images (Rodt, Ratiu et al. 2002), Figure 4.2. The thresholding method is useful when the structures to be segmented show a large difference in the pixel intensity values. Its main limitations are that, in its simplest form, only two classes are generated, and it cannot be applied to multichannel images. In addition, thresholding typically does not take into account the spatial characteristics of an image. This causes it to be sensitive to noise and intensity inhomogeneities, which can occur in CT images. Both of these artifacts essentially corrupt the histogram of the image, make the separation more difficult (Pham, Xu et al. 2000).

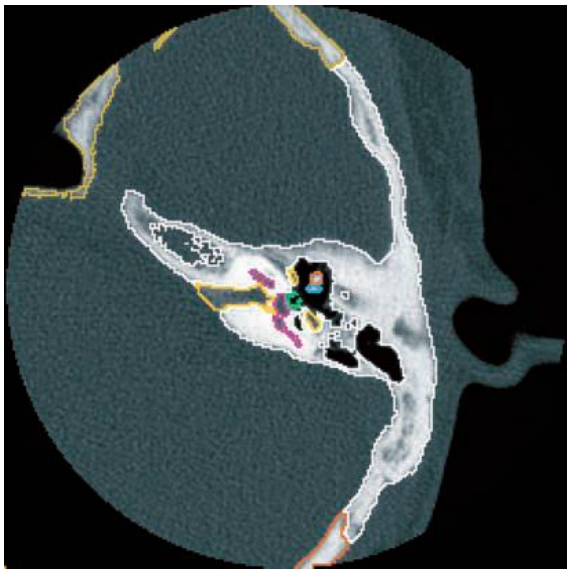


Figure 4.2: Axial CT image of the temporal bone and adjacent structures showing segmented structures in different colors (from (Rodt, Ratiu et al. 2002)).

Lee et al. (2010) introduces the 3D reconstruction through volume rendering and then the virtual endoscopy in the analysis of the middle ear, Figure 4.3. Methods for the 3D visualization of structures in medical images have been frequently used, and two of those methods are the volume rendering and the surface rendering (Kim, Kwon et al. 2002).

The volume rendering algorithm was proposed in the 1970's (Calhoun, Kuszyk et al. 1999) and it is a computer graphics technique whereby the object or phenomenon of interest is sampled or subdivided into a set of cubic building blocks, called voxels (or volume elements). A voxel is the 3D counterpart of the 2D pixel and is a measure of unit volume (Jun, Song et al. 2005). On the other hand, the surface rendering algorithm is an indirect method of obtaining an image from a volume dataset. The volumetric data

must first be converted into geometric primitives, by a process such as isosurfacing, isocontouring, surface extraction, or border following. These primitives are then rendered for display using conventional geometric rendering techniques. In contrast, a volume-rendering algorithm provides a method of directly displaying the data without any intermediate surface representations. Both algorithms have advantages and disadvantages.

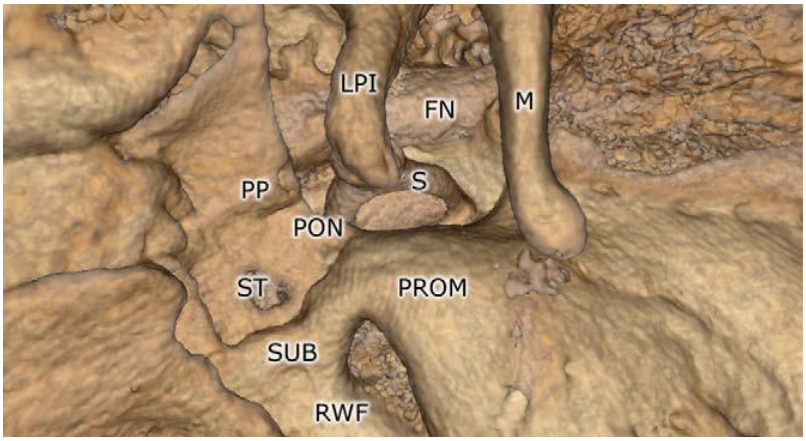


Figure 4.3: Rendering of the middle-ear structures using volume renderer: the facial nerve (FN), long process of the incus (LPI), malleus (M), ponticulus (PON), pyramidal process (PP), promontory (PROM), round-window fossa (RWF), stapes (S), sinus tympani (ST) and subiculum (SUB) (from (Lee, Chan et al. 2010))

A major advantage of the volume-rendering algorithm is that the 3D volume can be displayed without any knowledge of the geometry of the dataset and hence without intermediate conversion to a surface representation. This conversion step in a surface rendering algorithm can be sometimes be quite complex, especially if surfaces are not well defined, and can require various user interventions (such as manual contour tracing in segmentation). On the other hand, because the 3D dataset is reduced to a set of geometric primitives in a surface-rendering algorithm, this algorithm can provide fast display and manipulation of the 3D reconstructions built. However, a surface rendering algorithm has the following disadvantages: it discards the interior of the objects and maintains the objects' shells; does not facilitate real-world operations such as cutting, slicing, or dissection; and does not enable artificial viewing modes, such as semi-transparency. By contrast, since the entire image stack data is used for a volume-rendering algorithm, computes using considerable memory and processing power are required to handle volumes rendered in this manner. Because the entire dataset is preserved in a volume-rendering algorithm, any part, including internal structures and details (which may be lost when reducing to geometric structures with surface rendering) may be viewed.

For medical purposes, a volume rendering algorithm is preferred because of its ability to contain data on the internal architecture, to give a texture for the density changes in the data, to allow easy and natural exploration of volumetric data, and to exclude the need for classifying or segmenting the data (Calhoun, Kuszyk et al. 1999; Seemann, Seemann et al. 1999; Kim, Kwon et al. 2002; Jun, Song et al. 2005; Lee, Chan et al. 2010). The 3D images of the middle ear are important for: the visualization of ossicles, the achievement of clinical diagnosis, the planning of surgery, as well as for the post-surgical treatment. Major reasons cited for reduced clinical usefulness of 3D visualization include poor scan quality and difficulty in navigating through the 3D environment (Lee, Chan et al. 2010). In addition, volume rendering algorithms are also used for to 3D reconstruction of the inner ear (Seemann, Seemann et al. 1999).

Apart from the volume rendering algorithm, for micro-CT images are used the Marching Cubes (MC) algorithm for to perform the 3D reconstruction of the middle ear and inner ear (Decraemer, Dirckx et al. 2003; Xianfen, Siping et al. 2005), Figures 4.4 and 4.5. The MC algorithm is a sequential-traversal method that was described for the first time in 1987 by Lorense and Cline (Vignoles, Donias et al. 2010).

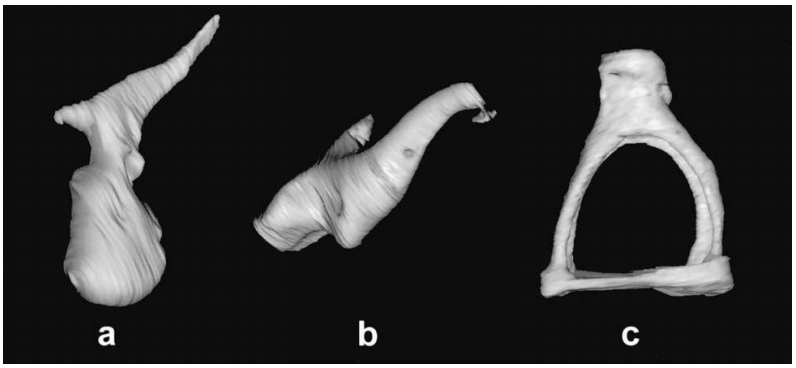


Figure 4.4: The 3D reconstructions of the three middle ear ossicles (malleus, a; incus, b; stapes, c) after segmentation of the serial section stack of the human temporal bone (from (Decraemer, Dirckx et al. 2003))

The standard MC constructs a facetized isosurface by processing the data set in a sequential, cube-by-cube (scanline) manner (Newman and Yi 2006). Vignoles et al. (2010) proposed a triangulation method, called simplified marching cubes (SMC), which can be viewed as a simplified alternative to the MC algorithm. The major advantage of the SMC is that all computations can be made with integer arithmetic (Vignoles, Donias et al. 2010). On the other hand, one advantage of the cube-by-cube processing of standard MC is that each edge intersection location only needs to be computed once. The MC standard also presents disadvantages. The first difficulty is observed when the MC algorithm is extended to higher dimensions, because becomes difficult to determine the intersection topology look-up table. There is a second

difficulty for high-dimensional data that is intersection topology table size increases with the dimension. In case of high-degree isosurfaces, one disadvantage of standard MC isosurfaces is that they can exhibit visible faceting artifacts. Datasets with cells of large size (relative to the desired viewing resolution) and datasets with cells of variant sizes, such as finite element data sets, tend to exhibit more severe artifacts. Use of a higher-degree isosurface representation is one means to reduce these artifacts. A key advantage of higher-degree patches is that a smoother isosurface can be produced. However, parametric polynomial fitting presents a heavier computational burden than triangular mesh fitting. Lastly, one complication that limits the MC is non-crisp structure boundaries. An example is in some medical data where non-uniform signal response of certain structures makes their boundaries not uniformly distinct (Newman and Yi 2006).

For bone which has a high-contrast ration with respect to surrounding soft tissue and air in micro-CT images, contouring can be performed by semi-automated algorithms. The shrink-wrapping algorithm is used in (Sim and Puria 2008) to tridimensional semi-automatic segmentation of the middle ear bone structures contained in micro-CT images, Figure 4.6.

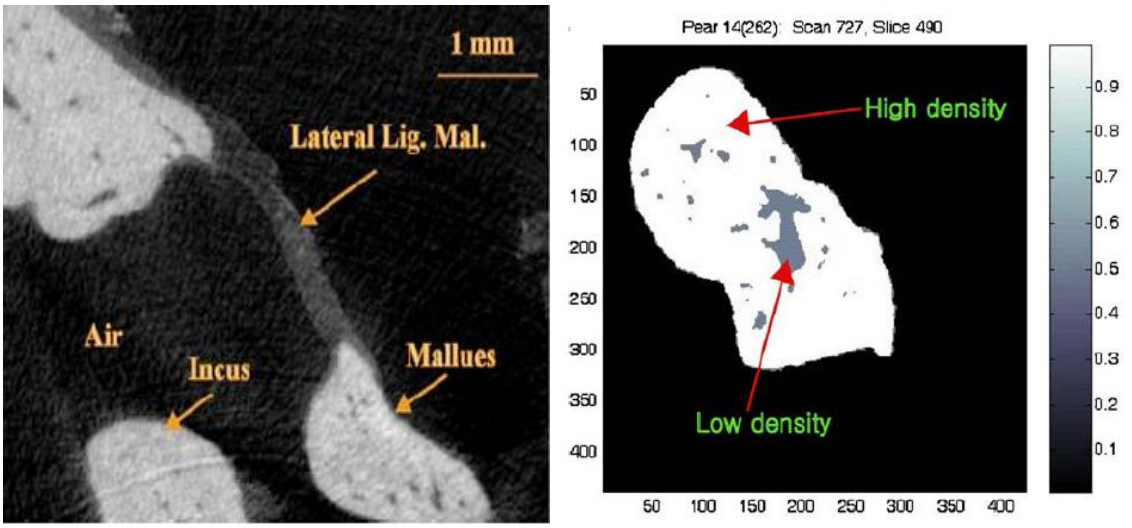


Figure 4.5: Micro-CT image (left) and segmented slice image of malleus (right) with a high-contrast ration and shrink-wrapping algorithm on slice 490 (from (Sim and Puria 2008)).

The shrink-wrapping algorithm was proposed in 1999 by Kobbelt et al., and can be characterized by a plastic membrane that is wrapped around an object that at the end of the process provides an exact imprint of the given geometry. The proposed approach is very similar to the concept of snakes or deformable surfaces known in image processing and computer vision. These techniques are used to extract contours from two and three

dimensional image data by optimizing the shape of a polygon or a triangle mesh (Kobbelt, Vorsatz et al. 1999). The shrink-wrapping algorithm is a shape reconstruction algorithm.

The shape reconstruction algorithms are characterized by two main stages: data acquisition and surface reconstruction. The data acquisition stage acquires an accurate, possibly unorganized, 3D point cloud from a real object, while the surface reconstruction stage converts the point cloud into a smooth surface (Koo, Choi et al. 2007). Furthermore, the shrink-wrapping algorithm is characterized by its high accuracy; however, there is a need of to perform the manual segmentation of the bone boundaries in the first image (Sim and Puria 2008). In addition, the algorithm is sensitive to artifacts appearance. The artifacts may result from noise presence as also from the lack of operator precision (Kobbelt, Vorsatz et al. 1999; Koo, Choi et al. 2007). The major advantage of the tridimensional structure bounded by surfaces is the ease in calculating its dimensions and its volume and so with this algorithm can be supported clinic conclusions.

4.4 Segmentation of the Inner Ear

The cochlea is one of the inner ear structures and, according to Xianfen et al. (2005), the segmentation of the cochlea is generally performed by using manual adjustments after incomplete automatic segmentation algorithms, such as region growing, threshold and morphologic filtering. Additionally, the deformable active contour algorithm is used for the segmentation of the cochlea by Yoo et al. (2001). Snake model has been applied successfully to a variety of image segmentation tasks, but it usually requires user interaction to locate the initial contour and adjust internal parameters of the model.

The basic knowledge to be used in cochlear segmentation is that the outer surface of the human cochlea is fairly smooth. The snake segmentation algorithm is considered appropriate for this application, because it can control the smoothness and adjust the convergence property. However, the snake modeling requires human interactions, namely, adjustment of processing parameters and specification of an initial contour. It cannot be operated well in mixed regions, in which edges to distinguish the cochlea are weak, and the intensity distribution of mixed structures is similar to that of the cochlea. Furthermore, semiautomatic segmentation systems using direct visual feedback have been applied to various medical images to reduce the time-consuming manual user interaction (Yoo, Wang et al. 2001). Moreover, this method is based on the assumption

that there are no significant changes in the shape of the cochlea between adjacent CT image slides.

Xianfen et al. (2005) applied the 3D narrow band level set segmentation algorithm, which not only kicks over the traces but also saves much time of post-processing to segment the cochlea, Figure 4.6. However, the 3D level set algorithm is used rarely, especially for the segmentation of the inner ear cochlea, but the 3D level set algorithm is adopted by Xianfen et al. (2005) for the following reasons: First, arbitrarily complex shapes can be modeled and topological changes, such as merging and splitting, are handled implicitly during the evolution of the level set function. Second, relatively small features can be segmented with the level set model. Third, the level set algorithm can control the smoothness of the contour. Additionally, the inner ear and cochlea are small and complex anatomical structures with fairly smooth outer surface. Hence, the level set algorithm is considered appropriate for this application. In the final step of the 3D narrow band level set, the segmented results are rendered with the Marching Cubes algorithm. A negative aspect of this algorithm is the need of to perform manual segmentation to remove several closely interconnected regions in boundaries in each image slice. The level set methods are very computationally expensive since all mesh points need updating every time, but the narrow band level set is an efficient modification of the level set methods, working only in a neighborhood of the zero level set (Xianfen, Siping et al. 2005; Mille 2009).

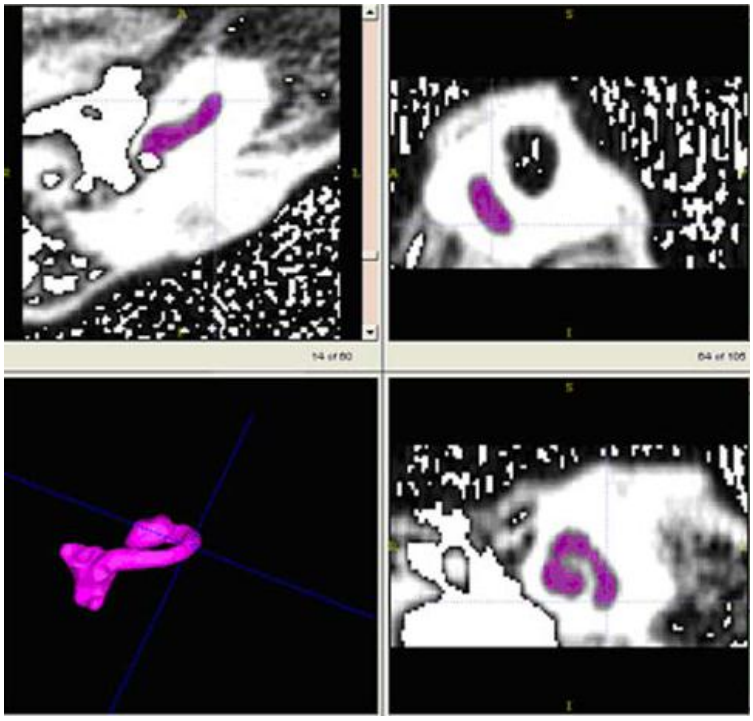


Figure 4.6: Correlation of points on the surface to the three orthogonal 2D sections using the marching cubes algorithm (3rd image) and cochlear segmentation in images 1st, 2nd, 4th performing

with a narrow band level set
(from (Xianfen, Siping et al. 2005)).

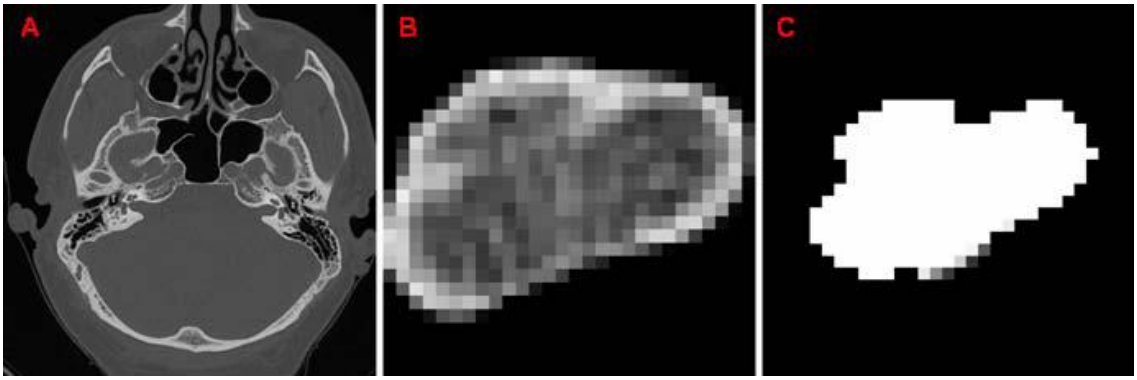


Figure 4.7: The result of applying a Connected Threshold region growing. From left to right: the original CT image containing the temporal bone (A), a stencil produced from the image containing the cochlea (B) and the result of the segmentation, which contains the pixels included in the region of interest (C) (from (Todd, Tarabichi et al. 2009)).

Todd et al. (2009) applied the connected threshold region growing algorithm for semi-automatic extraction of the cochlea from the spiral-CT images, Figure 4.7. Connected threshold region growing requires minimal user input, and its main criterion for inclusion of pixels in the region of interest (ROI) is a pixel intensity threshold range. The user specifies as input the index of a seed pixel and the lower and upper threshold limits. Pixels are included into the ROI if their intensity values are within the range specified. In order to iterate through pixels within the image and establish the ROI, the connected threshold applies a flood iterator for visiting neighboring pixels. Since the algorithm requires only three input values, it is ideal for applications that require a more automated approach, minimizing user input. One limitation of the region growing method is that segmentation results may depend on selection of the seed pixel; its location and intensity. Upper and lower threshold values used in Todd et al.(2009) were -800 and 900, respectively. Furthermore, connected threshold region growing has also been applied for segmentation of anatomical landmarks surrounding the cochlea; most importantly for implant cochlear pre-operative planning, the external ear canal. This structure provides the surgeon with crucial information for determining the most appropriate path to take on approach to the cochlea, for implant cochlear insertion (Todd, Tarabichi et al. 2009).

Poznyakovskiy et al. (2008) also used active contours (snakes) for segmenting the cochlea in micro computed tomography images, Figure 4.8. In this type of images is usually observed the presence of noise and partial volume effects. In order to suppress

artifacts on tomography images, a nonlinear smoothing method, anisotropic diffusion was applied.

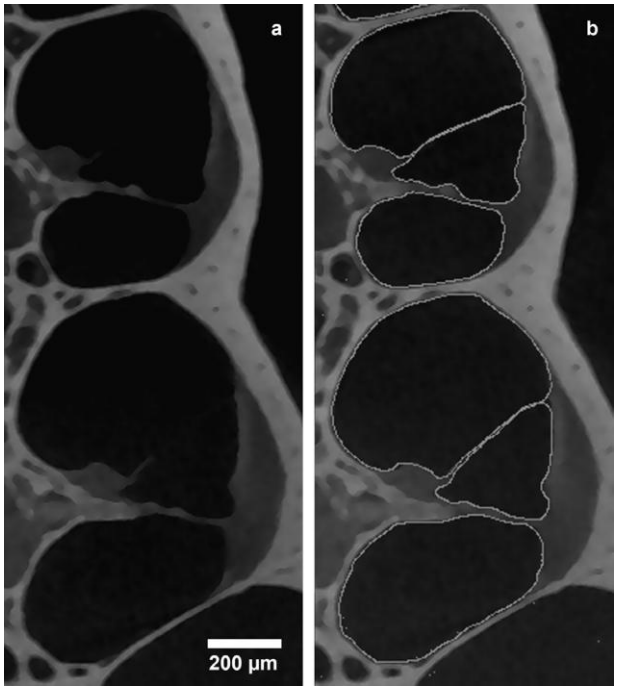


Figure 4.8: Segmentation with active contours: Image in cylindrical projection at 120º denoised with anisotropic diffusion (a); Image after segmentation with active contours (b) (from (Poznyakovskiy, Zahnert et al. 2008)).

The vestibular system is another structure of the inner ear, and it plays an important role in maintaining the equilibrium by sensing body movements and sending signals to the brain. Located in the inner ear compartment of the temporal bone, the vestibular system contains both osseous and membranous components. There are several challenges in developing such an automatic segmentation system for the vestibular system. Firstly, the diameter of the semicircular canals is as small as the magnitude of the voxel size. The cross-sectional diameter of the semicircular canal is around 0.8 mm in Humans (Gray 1985). Secondly, image intensities of the semicircular canals are not homogenous. Figure 4.9 shows the volume rendering via maximum intensity projection of a right-side vestibular system from MR images. Furthermore, the vestibular system is of similar image intensity as the inner ear, and is connected anatomically with the cochlear and auditory nerves. To analyze the vestibular system alone, the connected non-vestibular structures need to be removed and to enhance edges for better segmentation, MR images were processed by Shi (2010) using the anisotropic diffusion filter, which smoothed the image but preserved salient edges.

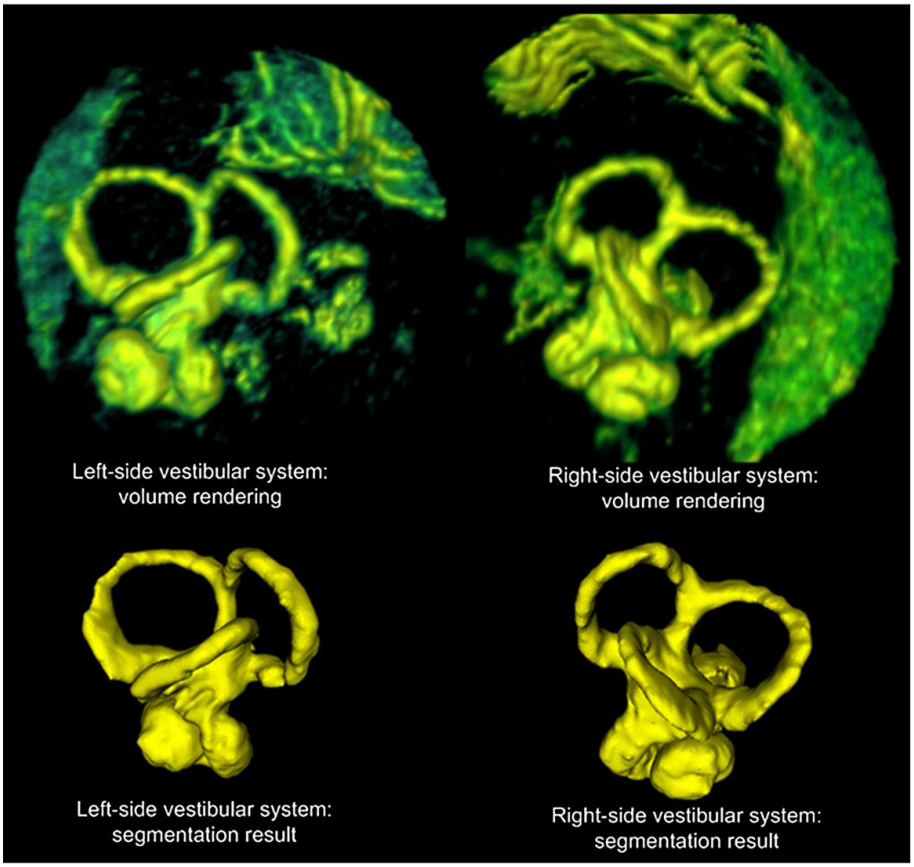


Figure 4.9: In top row is shown the volume rendering results within a ROI and in bottom row the corresponding segmentation results (from (Shi, Wang et al. 2010)).

Thereafter, Shi et al. (2010) applied the k-means clustering as a pre-segmentation step to categorize the voxels into background and foreground based on their signal intensities. The resultant cluster with smaller size was considered to be the foreground. The foreground cluster contained several connected components, among which the largest was chosen as a coarsely defined vestibular region. The surface of the coarsely segmented vestibular system was extracted using the marching cube algorithm and represented as a triangle mesh. Non-vestibular auditory structures, the cochlea and auditory nerves, are irrelevant to the function of balance control and should be excluded in the analysis of the vestibular system. In order to automatically remove them, Shi et al. (2010) defined a cutting plane on the reference vestibular surface of a healthy subject. By rigidly aligning the reference and the study surfaces using the iterative closest point (ICP) algorithm. To further refine the preliminary segmentation result, Shi et al. (2010) applied the deformable model method (McInerney and Terzopoulos 1996). In addition, in order to validate the segmentation results, Shi et al. (2010) performed the quality control over all the segmentation results by visually comparing the segmented surface with the volume data in the region of interest visualized using the volume rendering.

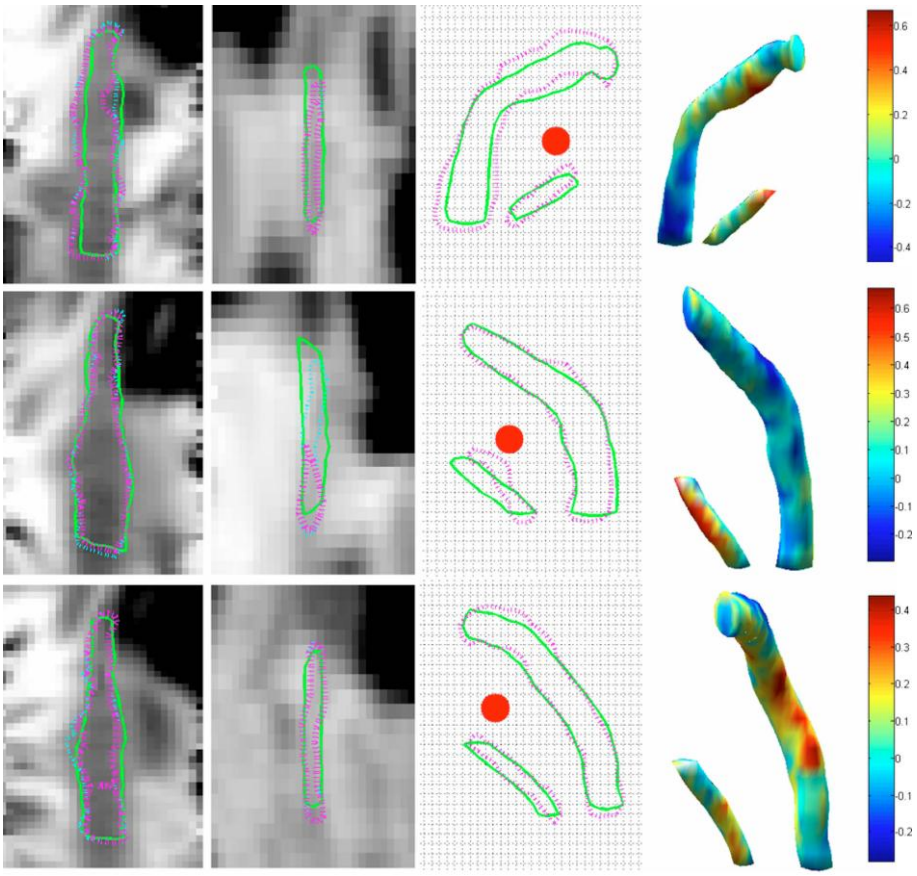


Figure 4.10: Segmentation results in CT volumes of the chorda tympani and facial nerve using the atlas based segmentation (from (Noble, Warren et al. 2008)).

Segmentation of the structures of the ear would improve trajectory planning safety and enable the possibility of automated planning. Two important structures of the ear, the facial nerve and the chorda tym pani, are difficult to segment with traditional methods because of their size (diameters as small as 1.0 and 3.0 mm, respectively), the lack of contrast with adjacent structures, and large interpatient variations. Atlas-based segmentation is a common technique, which relies on image registration, to perform automatic segmentation of general structures in medical images, and this technique was used by Noble et al. (2008) to segment the facial nerve and the chorda tympani, Figure 4.10. Noble et al. (2008) used a novel method that combines an atlas-based approach with a minimum cost path finding algorithm. The atlas is used to create a spatially varying cost function, which includes geometric information. Once the cost function is evaluated, a 3D minimum cost path finding is computed. This process is used to extract the centerline of the facial nerve and of the chorda. The centerlines are then expanded into the full structures using a level-set algorithm with a spatially varying speed function. Moreover, Noble (2008) presented the first method for the automatic

segmentation of two critical structures for cochlear implant surgery: the facial nerve and the chorda tympani. Because of the shape and size of these structures, and because of the lack of contrast between these and surrounding structures, accurate segmentation is challenging.

Figure 4.11: Transverse segmentation differences between the deformed patient left inner ear (black) and the atlas (white) (from (Christensen, He et al. 2003)).

The inner ear also can be called the labyrinth. Which contains the semicircular canals, otolith organs (sensory receptors for balance), and the cochlea (sensory receptors of hearing), as it was already referred. Christensen et al. (2003) described an automatic method for making precise measurements of the inner ear anatomy also using a deformable atlas of the human inner ear in the CT images, Figure 4.11. In this approach, the CT image of the inner ear anatomy from an individual is placed in a common orientation, and then the point-by-point correspondence between it and the deformable atlas is determined. In this method, the image data were converted to unsigned 8-bit format to reduce the storage using linear interpolation. Then, the images were thresholded using a threshold value of 134, which corresponds to the range from 1120 to 1135 Hounsfield units in the signed 16-bit intensity, to segment the image into regions of bone and nonbone. The threshold value of 134 was selected to only include the fluid component of the inner ear. The result of thresholding contains more than the objects of interest, which are connected with the objects through the fallopian canal, the opening of the internal auditory canal, the oval window, the round window, and openings on the semi-circular canals (Christensen, He et al. 2003). Furthermore, a deformable atlas was used to make automatic measurements by transforming its shape into the shape of a target or patient data set. All of the measurements were transformed through the mapping that deformed the atlas into the target shape, making them specific for the patient's data set. In other words, the segmentation corresponding to the cochlea was transforming from its location in the atlas to its location in the target data set. The volume of the patient's cochlea was then computed by summing up the voxels that were labeled by the deformed atlas, producing an automatic volume measurement of the patient's cochlea. Similarly, the landmarks in the atlas were mapped to their corresponding locations in the target data set. The distances between these new landmark locations were used to produce automatic measurements of the patient's inner ear anatomy (Christensen, He et al. 2003). As can be seen from the description above,

the automatic measurements produced by the deformed atlas are only as good as the image registration algorithm used to estimate the point-by-point correspondence.

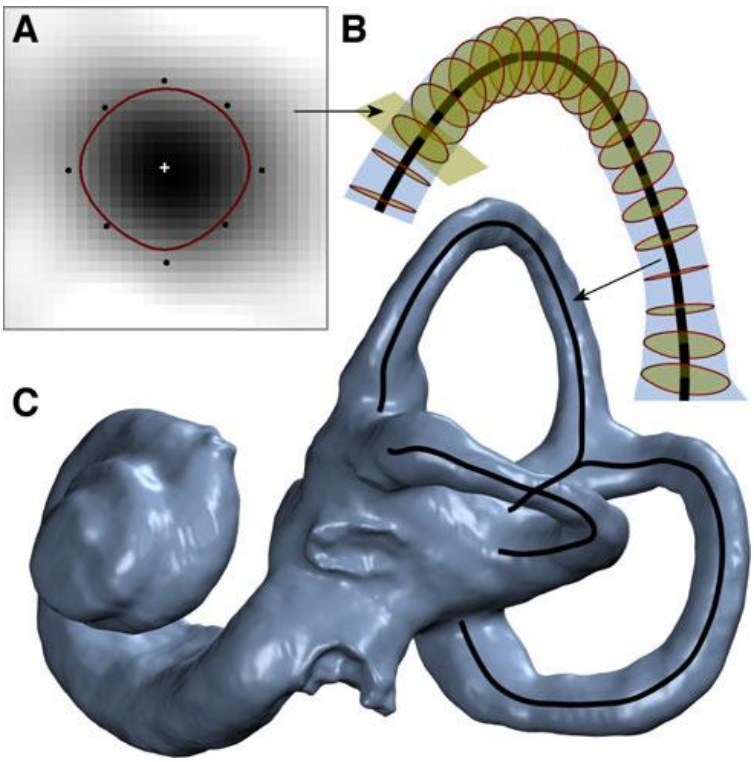


Figure 4.12: Modeling of SCCs: A cross-sectional slice of the bony canal is modeled using a B-spline active contour, with the centroid determined by a center of mass calculation (A). This modeling is performed along the entire length of the canal, with the contour centers tracing out the 3D geometrical centroid path (B). Modeling of the all three bony canals produces the complete canal centroid path shown overlaid on the labyrinth reconstruction (C) (from (Bradshaw, Curthoys et al. 2010)).

Also to perform the segmentation of the semicircular canals, but in this case in micro-CT images, Bradshaw et al. (2010) used the watershed algorithm, designed for boundary determination in situations where objects appear to overlap or are blurred together. The strategy used to reconstructing a complete SCC is by combining the cross-sectional modeling with an automated tracking system that propagates the active contour along the canal and automatically detects canal endpoints, Figure 4.12. The steps of this are: Firstly, an operator selects a starting position anywhere on the canal. Secondly, a 2D multiplanar reformatted CT slice of the canal cross-section is extracted from the 3D image at the designated position. Then, an active contour modeling is performed on the slice, reconstructing the canal cross-sectional morphology. In the next step, the centroid is calculated as the contour's center of mass and the position and orientation of the next slice is predicted from the previous centroids. Then, if a canal endpoint is not detected, is necessary to return to the second step.

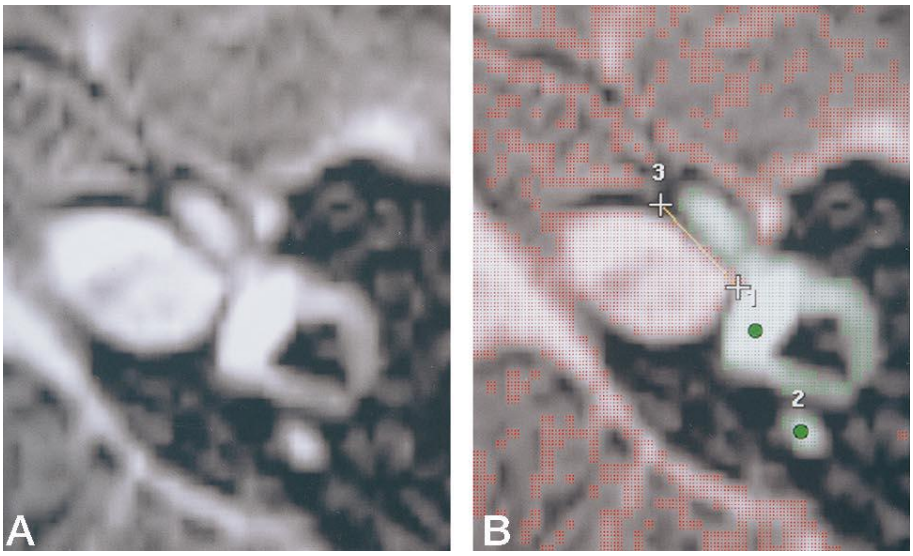


Figure 4.13: Axial T2-weighted MR image at the level of the vestibule (A). On-line segmentation (B) resulted in highlighting in red all pixels with signal intenstities above the set threshold. Pixels highlighted in green were those contiguous with the seed placed in the vestibule by the observers. Structures contiguous with but not part of the inner ear (internal auditory canal) were manually excluded (line 1-3) from the selected volume of interest (from (Melhem, Shakir et al. 1998)).

Finally, when the entire canal is traversed in both directions, the centroids are combined to form the 3D centroid path. The automatic detection of the canal endpoints uses the cross-sectional area to determine whether the contour has either passed into the vestibule or has reached the bifurcation of the common crus. The two boundary slices are adjusted via manually guided rotations to ensure a consistent alignment with the ends of the canal. Additionally, after the centroid path sections of the three canals and the common crus have been determined, the common cross section is attached to both the anterior and posterior canal sections. The joint is calculated using spline interpolation to maintain a smooth and continuous transition with the geometry at both ends. Each canal path is three-dimensionally resampled using spline interpolation to the predetermined number of samples and with a uniform sampling interval not exceeding 0.1 mm. The using of the active contours have several advantages, because they not critically depend upon subjective parameter selection as object boundaries are identified using image gradient information and they are less sensitive to artifacts such as noise and variations in background intensity (Bradshaw, Curthoys et al. 2010).

Finally, Melhem et al. (1998) performed the segmentation of the bone labyrinth in MR images using two algorithms; first, a thresholding algorithm was used and then the region growing algorithm, Figure 4.13.

4.5 Summary

In medical images of the ear have been applied several algorithms to segment and modeling the anatomic structures involved. Some of these algorithms are indicated in Table 3.1.

Table 4.1: Segmenting and Modeling methods that have been used in medical images of the ear.

Anatomic structure	Imaging Technique	Segmentation / 3D reconstruction Method	References
External Ear: Tympanic Membrane	Video Otoscopy	Generalized Gradient Vector Flow (GGVF snake)	(Xie, Mirmehdi et al. 2005)
	Otoscopy	Active Contours Segmentation Method of the Mumford-Shah	(Comunello, Wangenheim et al. 2009)
Middle Ear: Ossicles	Micro-CT	Threshold	(Lee, Chan et al. 2010)
	Magnetic Resonance	Threshold	(Rodt, Ratiu et al. 2002)
	Micro-CT	Volume Rendering	(Lee, Chan et al. 2010)
	Spiral-CT	Surface Rendering	(Seemann, Seemann et al. 1999)
	High Resolution Spiral-CT	Marching Cubes	(Jun, Song et al. 2005)
	Micro-CT	Shrink-Wrapping	(Sim and Puria 2008)
Inner Ear: Cochlea	Spiral-CT	Region Growing	(Xianfen, Siping et al. 2005)
	Spiral-CT	Morphologic Operations	(Xianfen, Siping et al. 2005)
	Spiral-CT	Threshold	(Xianfen, Siping et al. 2005)
	CT	Volume Rendering and Region Adaptive Snakes Modeling	(Yoo, Wang et al. 2001)
	Spiral-CT	3D Narrow Band Level Set and Marching Cubes	(Xianfen, Siping et al. 2005)
	Spiral-CT	Connected Threshold Region Growing	(Todd, Tarabichi et al. 2009)
	Micro-CT	Snakes	(Poznyakovskiy, Zahnert et al. 2008)
Inner Ear: Vestibular System	Magnetic Resonance	Clustering and Deformable models	(Shi, Wang et al. 2010)
Inner Ear: Facial nerve e timpani chorda	CT	Anatomic atlas and Minimum Cost Path Finding	(Noble, Warren et al. 2008)
Inner Ear: Semi-circular canals	Micro-CT	Active Contours	(Bradshaw, Curthoys et al. 2010)
Inner Ear: Bone labyrinth	Micro-CT	Watershed	(Chen and Chen 2009)
	Magnetic Resonance	Threshold	(Melhem, Shakir et al. 1998)
	CT	Threshold and deformable anatomic atlas	(Christensen, He et al. 2003)

Chapter V - Experimental Results

- Introduction;
- Study about Segmentation of Medical Images of the Ear;
- Preprocessing of the Medical Images;
- Application of Segmentation algorithms in the Medical Images of the Ear;
- Algorithm Selection;
- Summary

5.1 Introduction

In this chapter, it will present some results of the global topic of dissertation, in other words, some results of the segmentation structures of the inner ear.

The inner ear is present in the interior of the temporal bone. The structures of the inner ear that were segmented are the bony and membranous labyrinths. The first, the bony labyrinth, consists in the cochlea, semicircular canals and vestibule.

The medical images used in this work are in the DICOM format and they have a 512x512 size. The imaging technique used for obtaining these images is the Computer Tomography modality; consequently, the images have a grayscale color type. Furthermore, the techniques and algorithms are implemented in MATLAB.

The objective of this chapter is to present the results of the segmentation and it is to select a technique, which presents more advantages and the best results in the segmentation of the inner ear structures.

This chapter is organized into four sections. The first section is a study about segmentation techniques already used in medical images of the inner ear. In the second section a pre-processing technique for noise reduction and artifacts in the medical images is explored. The third section observes the application of segmentation algorithms in the medical images of the ear. Finally, in the last section is selected the best algorithm to segment these human structures is selected.

5.2 Study about Segmentation of Medical Images of the Ear

Through the Table 5.1 is possible to conclude that the type most commonly used in the medical imaging is the Computer Tomography.

This type of imaging technique provides 3D positional information and it offers excellent contrast for different tissues types. Furthermore, the computed tomography scans can be utilized even with the implant device in place. As a result, CT imaging has recently become a very capable tool in the research and development for improvement of cochlear implants. Clinically, CT patients scans are used in several treatment steps and before surgery, scans are commonly performed to identify abnormal anatomical structures and assist in surgery planning. After surgery, the patient is imaged to determine the position of the implanted array relative to ear anatomy, as a guide to selection of electrodes for stimulation programming and as an aid in interpreting the

reports of the percepts that patients experience. In these applications, the ability of the CT images to resolve submillimeter anatomical features and array electrodes is crucial. A limiting factor is the presence of image reconstruction distortions and artifacts in the vicinity of the electrode array, which in turn limits visualization of the fine anatomical detail near the electrode (Whiting, 2008).

Table 5.1: Segmentation methods and imaging techniques used for the ear anatomic structures analysis

Anatomic structure	Imaging Technique	Segmentation Method	References
External Ear: Tympanic Membrane	Video Otoscopy	Generalized Gradient Vector Flow (GGVF snake)	(Xie, Mirmehdi et al. 2005)
	Otoscopy	Active Contours Segmentation Method of the Mumford-Shah	(Comunello, Wangenheim et al. 2009)
Middle Ear: Ossicles	Micro-CT	Threshold	(Lee, Chan et al. 2010)
	Magnetic Resonance	Threshold	(Rodt, Ratiu et al. 2002)
Inner Ear: Cochlea	Spiral-CT	Region Growing	(Xianfen, Siping et al. 2005)
	Spiral-CT	Morphologic Operations	(Xianfen, Siping et al. 2005)
	Spiral-CT	Threshold	(Xianfen, Siping et al. 2005)
	Spiral-CT	Connected Threshold Region Growing	(Todd, Tarabichi et al. 2009)
	Micro-CT	Snakes	(Poznyakovskiy, Zahnert et al. 2008)
Inner Ear: Vestibular System	Magnetic Resonance	Clustering and Deformable models	(Shi, Wang et al. 2010)
Inner Ear: Facial nerve e timpani chorda	CT	Anatomic atlas and Minimum Cost Path Finding	(Noble, Warren et al. 2008)
Inner Ear: Semi-circular canals	Micro-CT	Active Contours	(Bradshaw, Curthoys et al. 2010)
Inner Ear: Bone labyrinth	Micro-CT	Watershed	(Chen and Chen 2009)
	Magnetic Resonance	Threshold	(Melhem, Shakir et al. 1998)
	CT	Threshold and deformable anatomic atlas	(Christensen, He et al. 2003)

In summary, the CT scanning presents a large variety of advantages for the observation of normal and abnormal structures of the ear, as well as implants that are inserted in the structures of the inner ear.

The middle ear and the inner ear are the two parts of the ear that can be observed by a Computed Tomography scan, but the middle ear modeling has proven to be quite accurate and vivid when compared to the actual morphology of the middle ear (Gentil,

2011; Liu, 2007). This way, the study and segmentation are only realized on the structures of the inner ear. Moreover, the most responsible organs for the hearing and balance are located in the inner ear.

Table 5.2: Segmentation Methods used in the inner ear anatomic structures when are only considered Computer Tomography images

Anatomic structure	Imaging Technique	Segmentation Method	References
Inner Ear: Cochlea	Spiral-CT	Region Growing	(Xianfen, Siping et al. 2005)
	Spiral-CT	Morphologic Operations	(Xianfen, Siping et al. 2005)
	Spiral-CT	Threshold	(Xianfen, Siping et al. 2005)
	Spiral-CT	Connected Threshold Region Growing	(Todd, Tarabichi et al. 2009)
	Micro-CT	Snakes	(Poznyakovskiy, Zahnert et al. 2008)
Inner Ear: Facial nerve e timpani chorda	CT	Anatomic atlas and Minimum Cost Path Finding	(Noble, Warren et al. 2008)
Inner Ear: Semi-circular canals	Micro-CT	Active Contours	(Bradshaw, Curthoys et al. 2010)
Inner Ear: Bone labyrinth	Micro-CT	Watershed	(Chen and Chen 2009)
	CT	Threshold and deformable anatomic atlas	(Christensen, He et al. 2003)

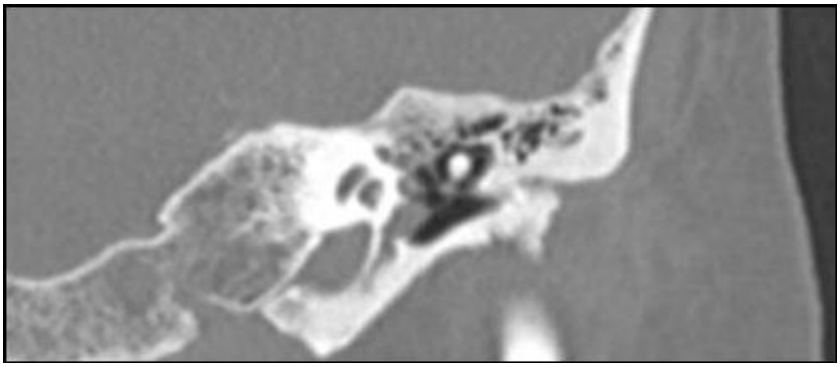
By analyzing the Table 5.2 is possible to see that the segmentation algorithms based on thresholding and on deformable models are often used in the inner ear Computed Tomography images. Therefore, for this work segmentation algorithms based in deformable models and thresholding were considered. These techniques were performed in Computed Tomography images, which can be seen in the Figure 5.1.

5.3 Preprocessing of Medical Images

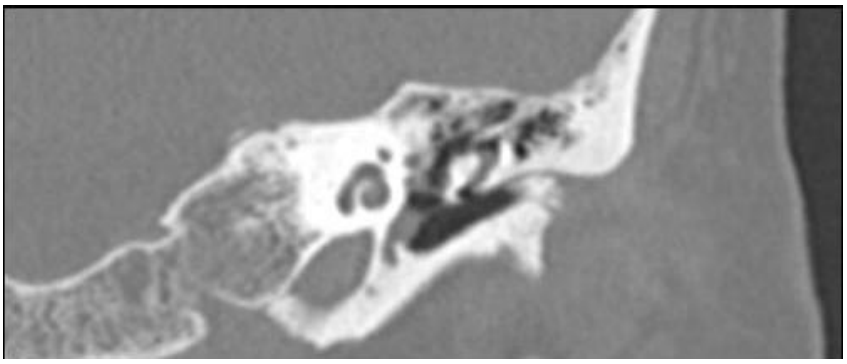
The techniques of image enhancement have a high technological potential and these are always used together with methods of analysis and quantification.

Enhancement techniques are procedures designed to manipulate an image in order to take advantage of the psychophysical aspects to the human visual system. Image enhancement is obtained with a variety of operations and the enhancement system has all the following functions:

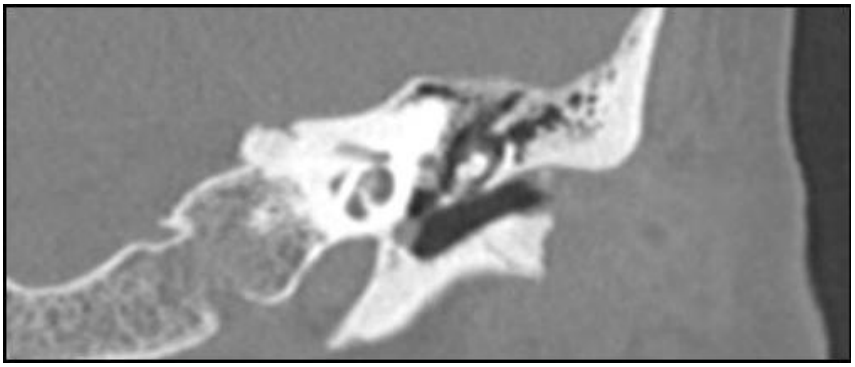
- Attenuate the effects of sub-sampling;



(a)



(b)



(c)

Figure 5.1: Sequence of three Computer Tomography imaging slices (slices: 10(a), 11(b), 12(c) of a total number of 34 slices of the left ear of a 68 years old female patient) of a temporal bone, in which is perfectly visible the inner ear structures

- Attenuate quantization effects;
- Remove noise and simultaneously preserve edges and image details;

- Avoid aliasing effects;
- Attenuate the blackness effect;
- Improve image contrast;
- Enhancement special features to be more easily detected by a machine or a human observer.

For medical images, noise is always involved in the signal due to the limitations of imaging hardware and protocols. Noise reduction is one of the most important objectives for medical image processing. Almost all model free segmentation methods are sensitive to noise.

For filter CT images, spatial filtering and nonlinear spatial filtering is used. The first type uses the spatial convolution and is characterized by a kernel (Gonzalez, 2004; Gonzalez, 2008), the Gaussian smoothing and the anisotropic diffusion algorithm are some examples of this filtering type (Feissel, 1984; Perona, 1990). In this type of images, Figure 5.1, the following methods are used:

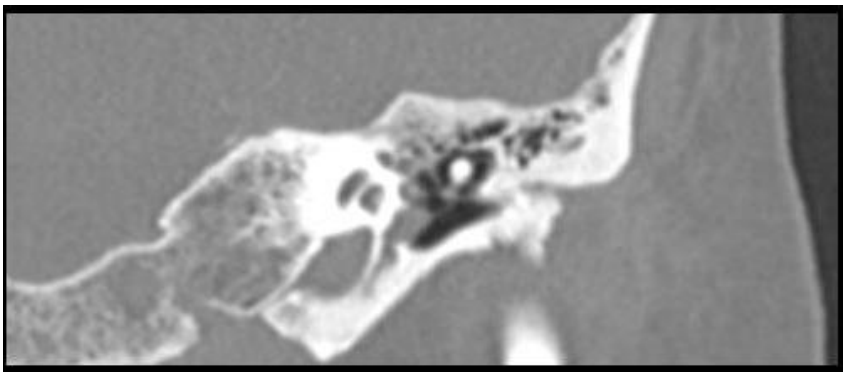
- Average filters;
- Gaussian filters;
- Anisotropic diffusion filters;
- Gradient operator;
- Histogram Equalization;
- Histogram Stretch.

The anisotropic diffusion filter is the most effective filter in the gray scale images (Xu, 1998; Poznyakovskiy, 2008; Tood, 2009), because this filter blurs areas of low contrast and enhances the high contrast (edges). Thus, the filter works as a high pass filter. However the noise and artifacts are characterized by a high frequency and because of this they can be considered like edges. Therefore, the using of a band-pass filter is the best choice.

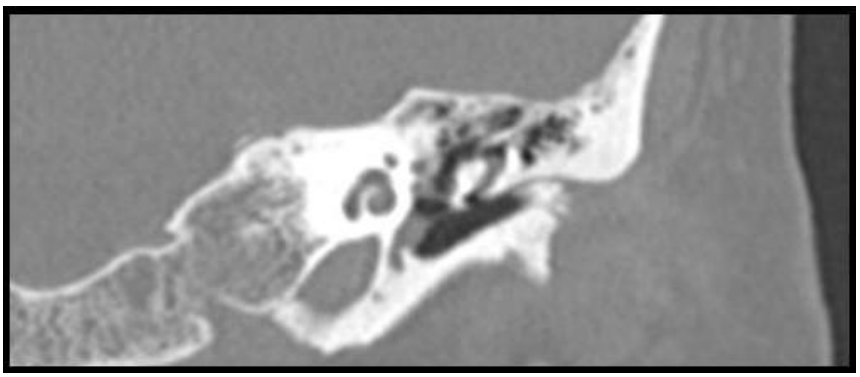
To perform the enhancement in the images of the Figure 5.1, I used three different functions of enhancement. First, I performed the histogram stretch on the image.

This operation is characterized by a gray-level scaling, when the higher and lower intensity, of the input image, is necessary to know.

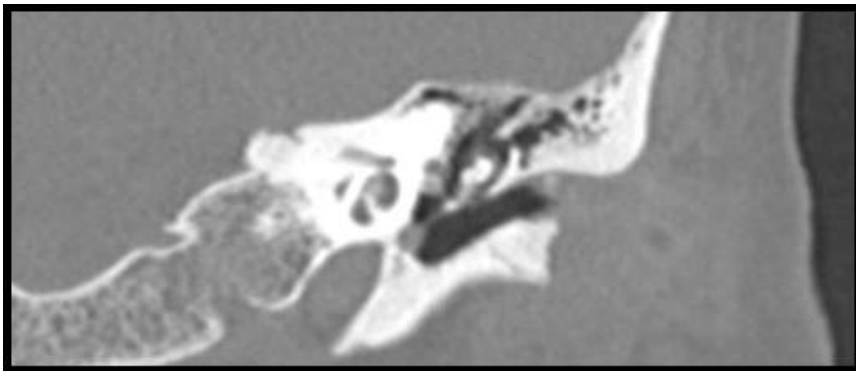
$$g = \frac{g_{\max} - g_{\min}}{f_{\max} - f_{\min}}(f - f_{\min}) + g_{\min} \tag{5.1}$$



(a)



(b)

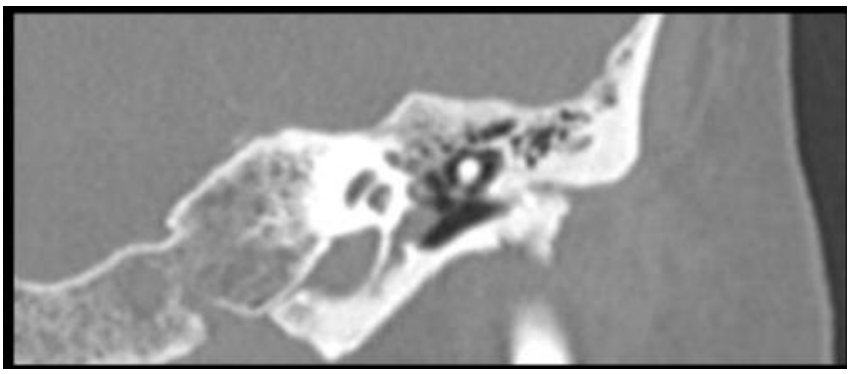


(c)

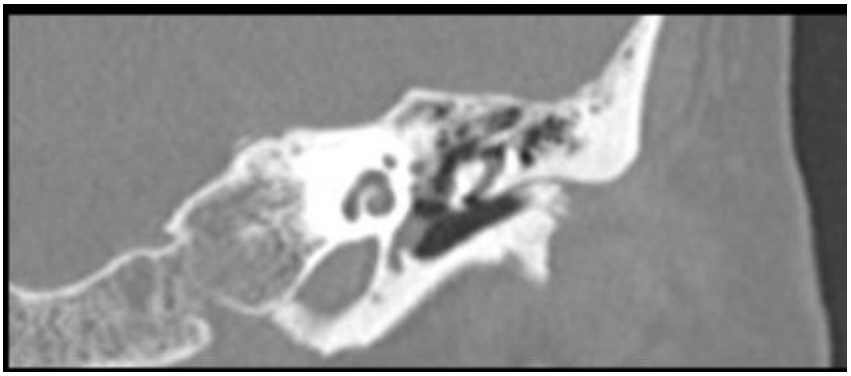
Figure 5.2: Results of the histogram stretch operation in the three slices

In the Equation 5.1 f represents the input image, $g_{\max}$ is the maximum value of the histogram result, $g_{\min}$ corresponds to minimum value of the histogram and $f_{\min}$ and $f_{\max}$ are respectively the intensity values minimum and maximum.

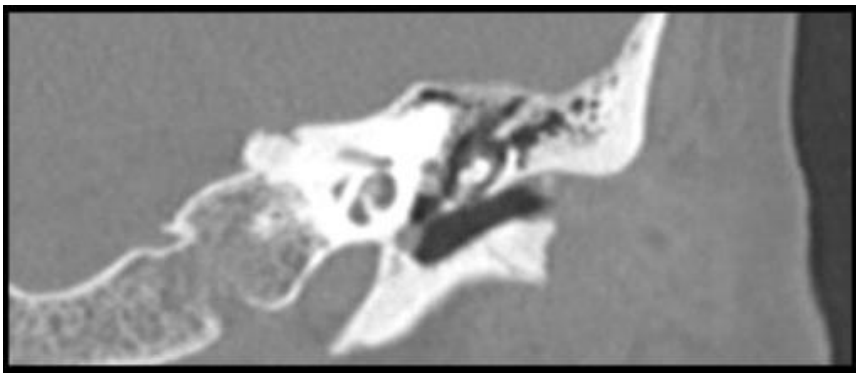
The histogram stretch was made to increase the contrast of the images and because of this contrast enhancement of the different tissues are more easily identified, Figure 5.2.



(a)



(b)

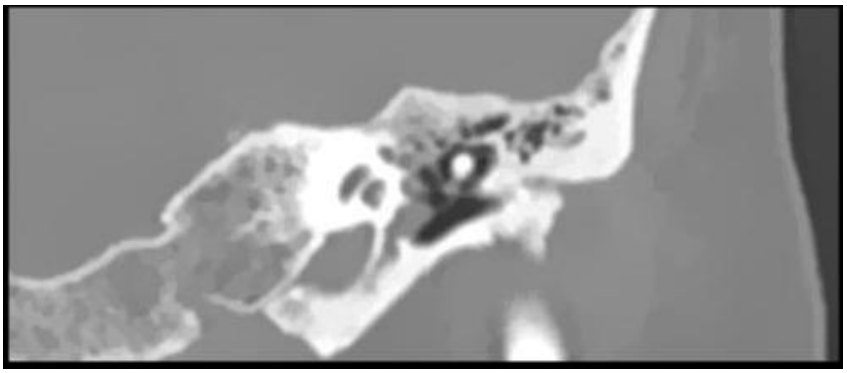


(c)

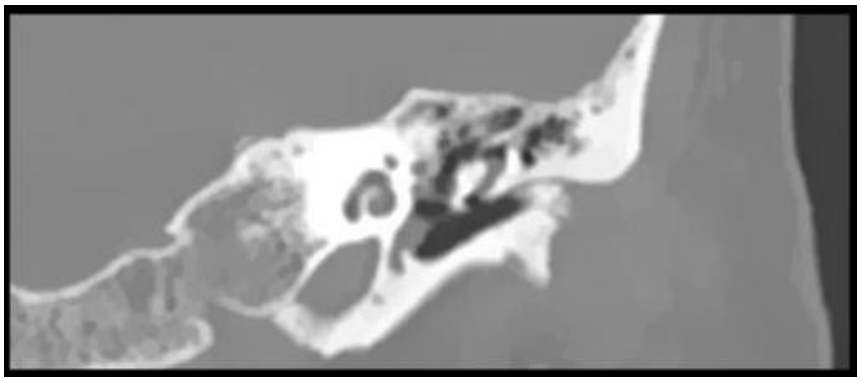
Figure 5.3: Results follow from the passing of the Gaussian filter in the images (a)-(b) of the Figure 5.2

The second step of the of the enhancement process was the Gaussian filtering with a 12x12 kernel and a standard deviation of 0.5. The Gaussian smoothing method tends to blur the sharp boundaries in the image while removing the noise, Figure 5.3.

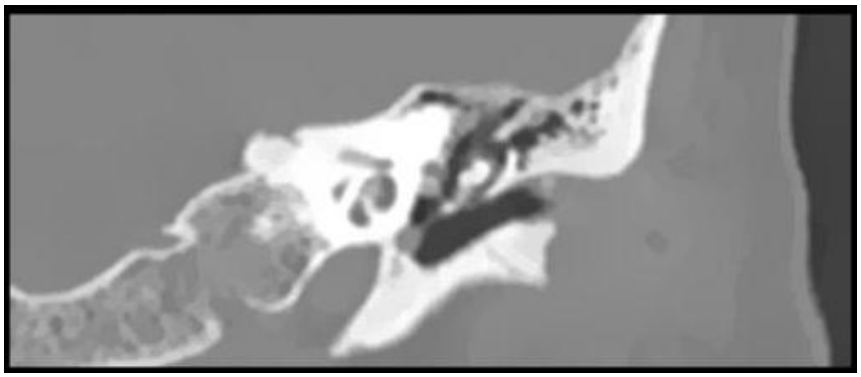
Finally, the third step of the enhancement process was completed with an anisotropic diffusion filter, Figure 5.4. Therefore, there was the use of a contrast increase operator and a band-pass filter.



(a)



(b)



(c)

Figure 5.4: Results of the anisotropic diffusion filter in the images (a)-(c) of the Figure 5.3

To reduce time consumption and the computational cost, the region of the inner ear was selected, Figure 5.5. The inner ear region is represented by the high intensity of the pixels. In this high intensity the membranous labyrinth is localized and from the

membranous labyrinth it is possible to create a surface that represents the boundaries of the temporal bone, this way the image size is reduced and the structures of interest are fully inserted into the image of work, in other words, in the region of interest.

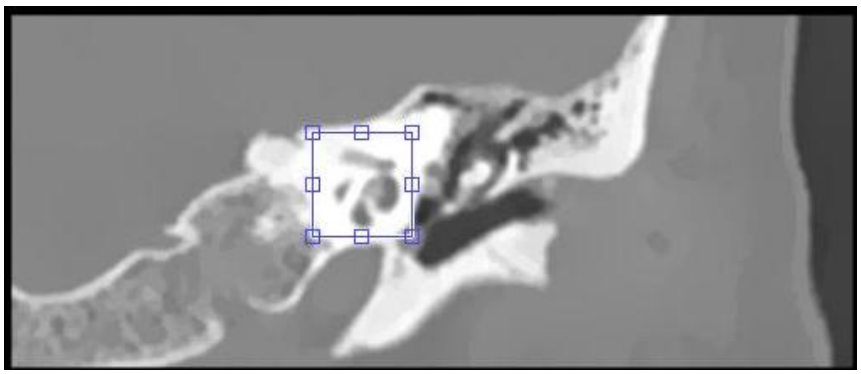


Figure 5.5: Region of interest selected in the figure 5.4 – (c)

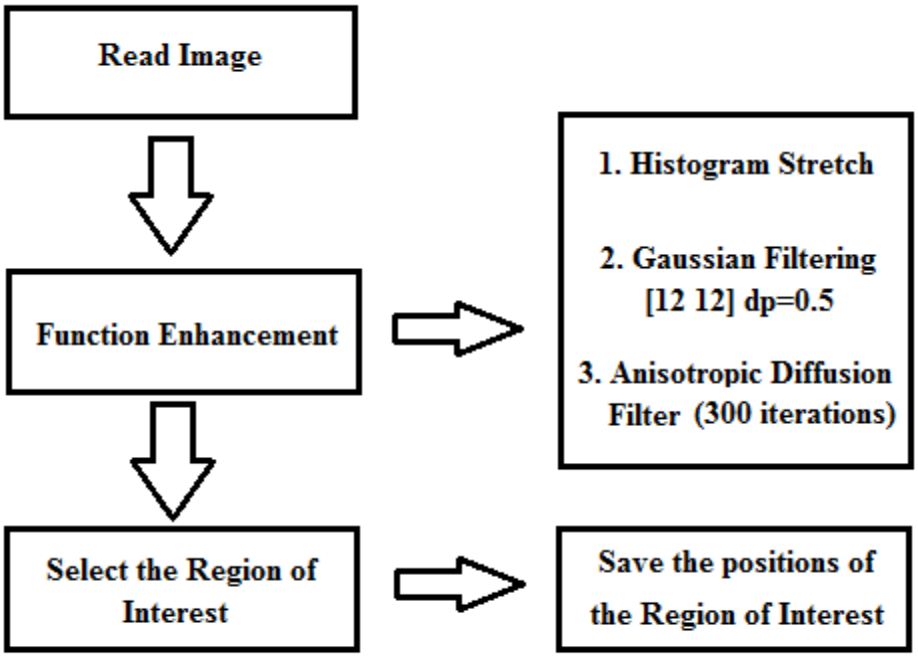


Figure 5.6: Method of preprocessing scheme

The preprocessing technique is sketched in the Figure 5.6 and the resulting images are exposed in the Figure 5.7. Furthermore, the resulting images are considered according to the location / position of the region of interest selected in the Figure 5.5, where, in this case, the four vertices of the rectangle region have the following coordinates (p_{xi}, p_{yi}), (p_{xi}, p_{yf}), (p_{xf}, p_{yi}) and (p_{xf}, p_{yf}), on which $p_{xi} = 185$, $p_{xf} = 241$, $p_{yi} = 223$ and $p_{yf} = 283$.

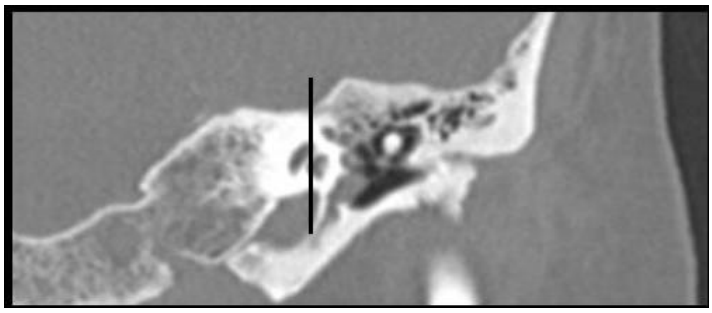
(a)

(b)

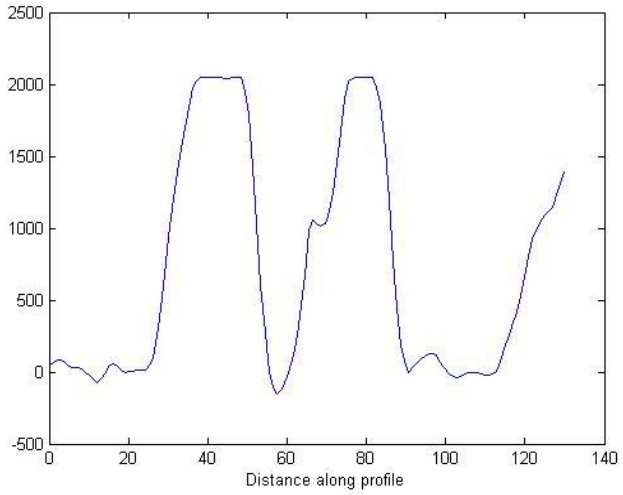
(c)

Figure 5.7: Resulting images of the preprocessing performed on the Figure 5.1 (a)-(c) and it is considered the rectangle region observed in the Figure 5.5

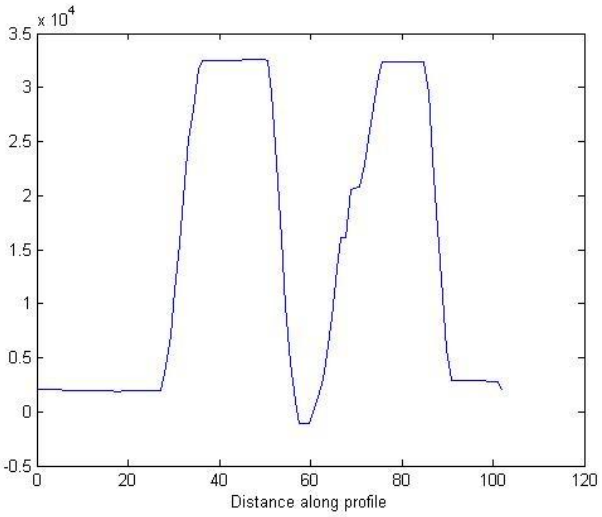
The effectiveness of the preprocessing method is verified through the analysis of a line, which has an edge inserted, of the original and enhanced image. However, the region of interest isn't considered because this stage is only responsible for the computational cost and processing time reduction.



(a)



(b)



(c)

Figure 5.8: The black line (a) was analyzed in the original image, Figure 5.1.a, and in the enhanced image, Figure 5.4.a. It is observed the intensity profile of the line black in the original image (b) and in the enhanced image (c)

Through the analysis of the Figure 5.8.b it is possible to observe that between the [0-20] and [90-100] the noise that was removed, Figure 5.8.c, by the function enhancement of the created preprocessing method. In addition, when the enhanced image was analysed and by comparing the two profiles, I immediately observed change of the pixel intensities, in the locations of edges, unlike of the intensity profile of the original image.

Therefore, the preprocessing method is effective in removing noise and it enables the decrease of the time consuption and computational cost.

5.4 Application of Segmentation algorithms in the Medical Images of the Ear

In this section, the segmentation methods used to acquire the inner ear structures of the images represented in Figure 5.4, as well as, the results of the tested segmentation algorithms is described.

Once verified that the algorithms based on thresholding technique and on deformable models technique are the most widely used in CT images of the ear, I denote that these two types will be analyzed. The analysis of the segmentation algorithms will be done from the observation of the segmentation results and by inspection of the advantages and disadvantages of each one.

This section is subdivided into three secondary sections. In the first, the segmentation algorithms are applied and the segmentation results are observed.. In the second section, the problems and the advantages of the segmentation algorithms performed are exhibited. Finally, in the third and last section the best method to realize the segmentation of the inner ear structures is selected.

5.4.1 Results

This subsection is dedicated to presenting the results of the following segmentation techniques:

- Otsu Method;
- Region Growing;
- Canny Detector;
- Watershed;
- Snake;
- Chan-Vese Model;
- Level set algorithm of the Li and Xu.

Otsu Method

Otsu method was proposed in 1979. This method is deduced by least square (LS) method based on gray histogram. Supposed that $G = [0, L-1]$ is the range of grayscale of image $f(x, y)$ and P_i is the probability of every grayscale, and the threshold value t has splitted the image in two classes which are $C_0 = [0, t]$ and $C_1 = [t+1, L-1]$. The two classes probability are $\alpha_0 = \sum_{i=0}^{t} P_i$ and $\alpha_1 = 1 - \alpha_0$ respectively. The average gray value of the two classes are $\mu_0 = \sum_{i=0}^{t} \frac{iP_i}{\alpha_0} = \frac{\mu_t}{\alpha_0}$ and $\mu_1 = \sum_{i=t+1}^{L-1} \frac{iP_i}{\alpha_1} = \frac{\mu - \mu_t}{1-\alpha_0}$ respectively, there in $\mu = \sum_{i=0}^{L-1} iP_i$, $\mu_t = \sum_{i=0}^{t} iP_i$. The criterion function has been defined as variance between the two classes, expressed as:

$$\eta^2(t) = \alpha_0(\mu_0 - \mu)^2 + \alpha_1(\mu_1 - \mu)^2 = \alpha_0\alpha_1(\mu_0 - \mu_1)^2 \tag{5.2}$$

Calculating the Eq.(5.2) above, I can obtain the maximum t which is the threshold value namely, and mark it t (Fang, 2009).

Considering the image results from the preprocessing stage, Figure 5.7, the code was performed on them which can be observed in the Table 5.3 and using it leads to obtaining the observed results in Figure 5.9.

In segmentation algorithms the objects of interest have one's value of the intensity. Once the membranous structures of the inner ear have low intensity it is necessary to reverse the pixel intensity, this is seen in the third line of the Table 5.3. At a previous stage the normalization of the pixels was realized. This process is shown in the first and second line of the Table 5.3. The last line of the same table represents an estimating of the threshold value by Otsu method and, then, it computed the binarization.

Table 5.3: Sequence of code for segment the inner ear structures with a Otsu Method

% Normalize
max_val = 1;
min_val = 0;
img_norm = normalize_img(roi, max_val, min_val);

%Normalize Function
function img_norm = normalize_img(img, max_val, min_val)
img_min = min(img(:));
img_max = max(img(:));
img_norm1 = (max_val - min_val)/(img_max - img_min);

```
img_norm = img_norm1 .* (img - img_min);
end
```

```
%Reverse intensity Values
img_norm_1 = imadjust(img_norm,[0;1],[1;0]);
```

```
% Otsu Method
level = graythresh(img_norm);
BW = im2bw(img_norm_1,level);
```

(a) (b) (c)

Figure 5.9: Results of the Otsu Method in the Figure 5.7

Region Growing

The region growing algorithm is a method that is included in the region-based algorithms, and, the region-based algorithms are inserted in the thresholding-based algorithms. Its idea comes from the observation that quantifiable features inside a structure tend to be homogeneous. Therefore, algorithms aim to search for the pixels with similar features values.

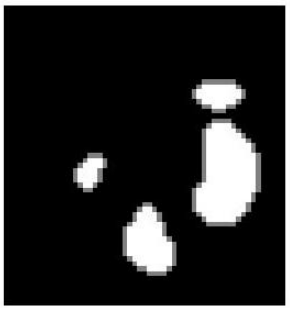
The region growing, in this case, starts with two or three seeds. This specific number of seeds is relative to the number of objects in the region of interest. After choosing the number and location of the seed points, the next step is to examine the neighboring pixels, so they are analyzed once at time and added to the region growing, if they are sufficiently similar based on a uniformity test, the procedure continues until no more pixels can be added. The result is then represented by all pixels that have been accepted during the region growing procedure. The uniformity test can be compared with the difference between the pixel intensity value and the mean intensity value, and when the difference is less than a predefined value the pixel is included in the region, otherwise, it is defined as an edge pixel. This process is explained in the Figure 5.10 and in the table 5.4.

In the Figure 5.11 the results of the region growing algorithms is illustrated. For each one, Figure 5.11.a-c, specific local seeds were used and the threshold limit value considered is between 0.2 and 0.8.

The Table 5.4 represents the MATLAB code, which was used to obtain the result observed in the Figure 5.11.c.

(a)

(b)



(c)

Figure 5.10: Obtained results from the region growing algorithms

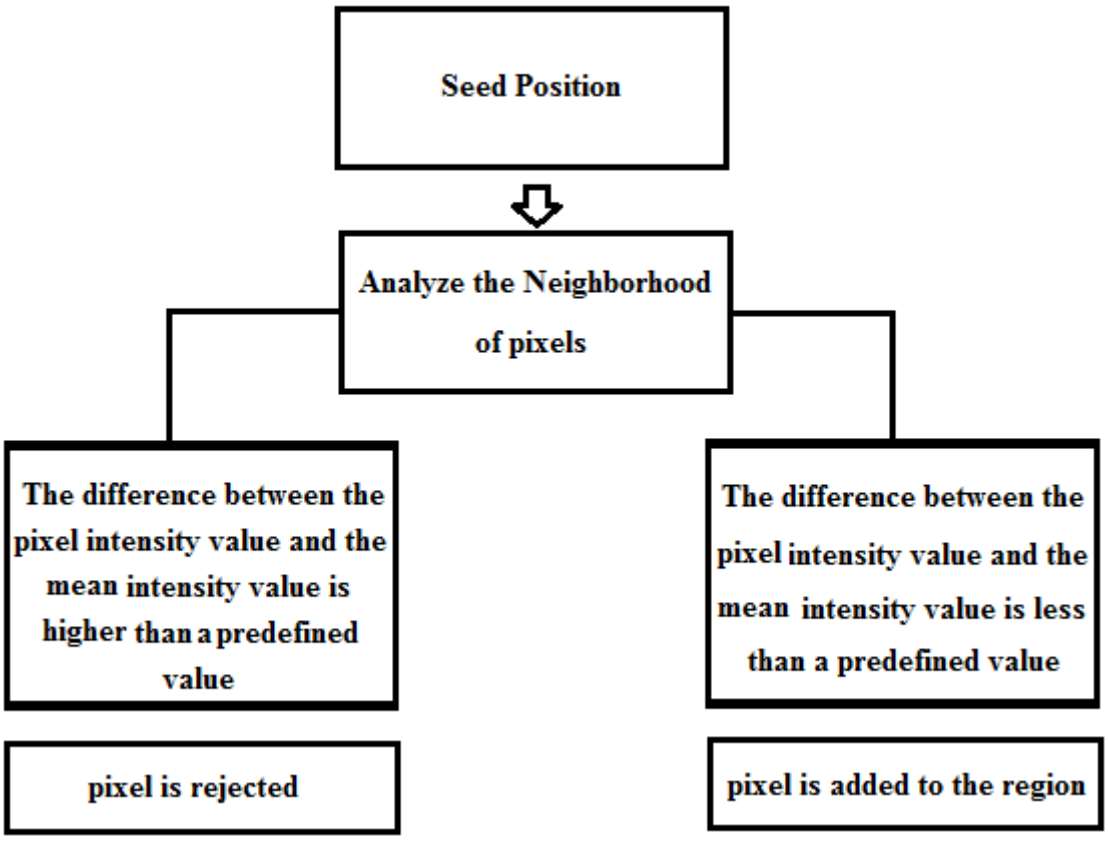


Figure 5.11: Representative scheme of the region growing algorithm

Table 5.4: Matlab code performed to obtain the Figure 5.11.c (Kroon, 2008; Kroon, 2009)

```
% Seed Local
x1 = 35;
y1 = 48;
x2 = 50;
y2 = 31;
x3 = 35;
y3 = 18;
x4 = 18;
y4 = 46;
```

```
%Normalization
max_val = 1;
min_val = 0;
img_norm = normalize_img(roi_2, max_val, min_val);
```

```
% Region Growing
J = regiongrowing(img_norm,x1,y1,0.18);

J1 = regiongrowing(img_norm,x2,y2,0.18);

J2 = regiongrowing(img_norm,x3,y3,0.08);

J3 = regiongrowing(img_norm,x4,y4,0.08);

RG = J + J1 + J2 + J3;
```

```
function J=regiongrowing(I,x,y,reg_maxdist)
% This function performs "region growing" in an image from a specified
% seedpoint (x,y)
%
% J = regiongrowing(I,x,y,t)
%
% I : input image
% J : logical output image of region
% x,y : the position of the seedpoint (if not given uses function getpts)
% t : maximum intensity distance (defaults to 0.2)
%
% The region is iteratively grown by comparing all unallocated neighbouring pixels to
the region.
% The difference between a pixel's intensity value and the region's mean,
% is used as a measure of similarity. The pixel with the smallest difference
% measured this way is allocated to the respective region.
% This process stops when the intensity difference between region mean and
% new pixel become larger than a certain treshold (t)
```

```
if(exist('reg_maxdist','var')==0), reg_maxdist=0.2; end
if(exist('y','var')==0), figure, imshow(I,[]); [y,x]=getpts; y=round(y(1)); x=round(x(1));
end

J = zeros(size(I)); % Output
Isizes = size(I); % Dimensions of input image

reg_mean = I(x,y); % The mean of the segmented region
reg_size = 1; % Number of pixels in region

% Free memory to store neighbours of the (segmented) region
neg_free = 10000; neg_pos=0;
neg_list = zeros(neg_free,3);

pixdist=0; % Distance of the region newest pixel to the regio mean

% Neighbor locations (footprint)
neigb=[-1 0; 1 0; 0 -1;0 1];

% Start regiogrowing until distance between regio and posible new pixels become
% higher than a certain treshold
while(pixdist<reg_maxdist&&reg_size<numel(I))

% Add new neighbors pixels
for j=1:4,
% Calculate the neighbour coordinate
xn = x +neigb(j,1);
yn = y +neigb(j,2);

% Check if neighbour is inside or outside the image
ins=(xn>=1)&&(yn>=1)&&(xn<=Isizes(1))&&(yn<=Isizes(2));

% Add neighbor if inside and not already part of the segmented area
if(ins&&(J(xn,yn)==0))
neg_pos = neg_pos+1;
neg_list(neg_pos,:) = [xn yn I(xn,yn)]; J(xn,yn)=1;
end
end

% Add a new block of free memory
if(neg_pos+10>neg_free), neg_free=neg_free+10000;
neg_list((neg_pos+1):neg_free,:)=0; end

% Add pixel with intensity nearest to the mean of the region, to the region
dist = abs(neg_list(1:neg_pos,3)-reg_mean);
[pixdist, index] = min(dist);
J(x,y)=2; reg_size=reg_size+1;

% Calculate the new mean of the region
```

```
reg_mean= (reg_mean*reg_size + neg_list(index,3))/(reg_size+1);

% Save the x and y coordinates of the pixel (for the neighbour add proccess)
x = neg_list(index,1); y = neg_list(index,2);

% Remove the pixel from the neighbour (check) list
neg_list(index,:)=neg_list(neg_pos,:); neg_pos=neg_pos-1;
end

% Return the segmented area as logical matrix
```

Canny Operator

The Canny operator closely approximates the operator that optimizes the product of signal to noise ratio and localization. Let $S[i, j]$ be the result of convolving the input image $I[i, j]$, with a Gaussian smoothing filter given by:

$$S[i, j] = G[i, j] * I[i, j] \tag{5.3}$$

The gradient of $S[i, j]$ is computed to produce two arrays $P[i, j]$ and $Q[i, j]$ for the x and y partial derivates:

$$P[i, j] \approx (S[i, j+1] - S[i, j] + S[i+1, j+1] - S[i+1, j]) / 2 \tag{5.4}$$

$$Q[i, j] \approx (S[i, j] - S[i+1, j] + S[i, j+1] - S[i+1, j+1]) / 2 \tag{5.5}$$

The magnitude and orientation of the gradient can be computed from:

$M[i, j] = \sqrt{P[i, j]^2 + Q[i, j]^2}$ and $\theta[i, j] = \arctan(Q[i, j], P[i, j])$, respectively.

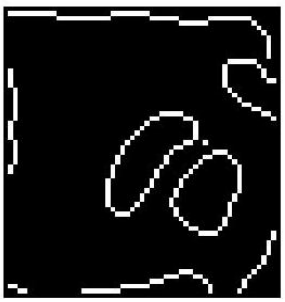
Through the table 5.5 is possible to see the method used for obtain the presented results in the Figure 5.12 and the method is based in the partial derivates described above.

Table 5.5: Procedure used for obtainning the results of the Figure 5.12

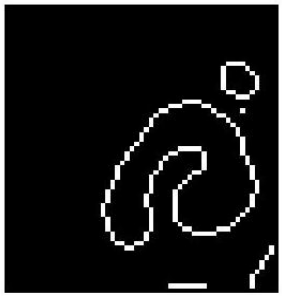
```
% Normalization
max_val = 1;
min_val = 0;
```

```
img_norm = normalize_img(roi, max_val, min_val);
```

```
% Canny operator
I_can = edge(img_norm, 'canny');
```



(a)



(b)

(c)

Figure 5.12: Resulted Image of the Canny operator

Watershed

The watershed method is classified into rain falling and water immersion. The watershed method used the prior information-based of the more frequently-used gradient function to improve the segmented results and employ the marker images as probes to explore a gradient space to unknown image and thus to determine the best matched object.

Watershed requires selection of at least one marker (seed point) interior to each object of the image, including the background as a separate object. The markers are chosen by an operator and the markers are provided by an automatic procedure that takes into account be application-specific knowledge of the objects.

Watershed method can be thought as a surface where the bright pixels represent mountaintops and the dark pixels valleys. The surface is punctured in some of valleys, and then slowly submerged into a water bath. The water will pour in each puncture and start to fill the valleys. However, the water from different punctures is not allowed to mix, and therefore the dams need to be built at the points of first contact.

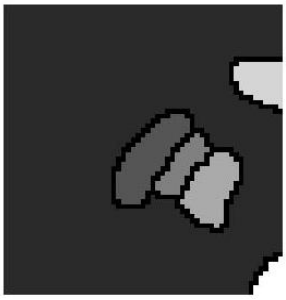
This algorithm is associated to the regional minimal set,

$$M = U_{i \in \Re} m_i \tag{5.6}$$

of an image f and can be defined as a complement of the union of all the retention basins $C_f(m_i)$:

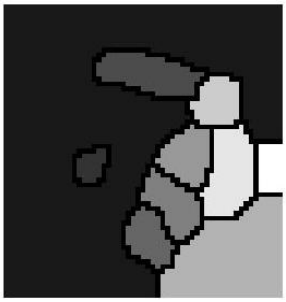
$$WL(f) = [U_{i \in \Re} C_f(m_i)]^c \tag{5.7}$$

For this algorithm is also shown the obtained result, Figure 5.13, by using the code shown in the Table 5.6.



(a)

(b)



(c)

Figure 5.13: Obtained results by using the watershed algorithm

In the Figure 5.13 a-c was used threshold values of 0.4, 0.1 and 0.075 respectively.

Table 5.6: Watershed algorithm represented with MATLAB code for obtain the result saw in the figure 5.13.c

```
%Normalization
max_val = 255;
min_val = 0;
img_norm = normalize_img(roi_2, max_val, min_val);
```

```
%Gaussian Filter
ImageScaled = mat2gray(img_norm);
sigma = 10;
hsize=(4*ceil(0.5)+1);
hg= fspecial('gaussian',hsize,sigma);
```

```
%Gradient
Image_out = imfilter(ImageScaled,hg);
```

```
h1 = fspecial('sobel');
h2 = h1';
Imageout = abs((imfilter(Image_out,h1).^2)+(imfilter(Image_out,h2).^2));
```

```
%Watershed
D = Imageout;
L1 = watershed(img_norm);
rgb1 = label2rgb(L1);
L2 = watershed(D>0.075);
rgb2 = label2rgb(L2);
```

Snake

The snake is a parametric deformable model and the main feature of it is to track the evolution through sampled contour points. The moving equation for the contour derived through energy functions and the basic premise of the energy minimizing-formulation of deformable contours is to find a parameterized curve that minimizes the weighted sum of internal and potential energies. The internal energy specifies the tension or smoothness of the contour and the potential is defined over the image domain and typically possesses local minima at the image intensity edges occurring at object boundaries. Minimizing the total energy yields internal forces and potential forces. The internal forces hold the curve together (elasticity forces) and keep it from bending too much (bending forces) while, external forces attract the curve toward the desired object boundaries.

A snake is a curve:

$$v(s) = (x(s), y(s)) \tag{5.8}$$

which mores through the spatial domain of an image to minimize the following energy functional:

$$E_{snake} = \int_{S=0}^{1} E_{\text{int}}(v(s)) + E_{image}(v(s))ds \tag{5.9}$$

$$E_{\text{int}} = \alpha(s)\left|\frac{d(v(s))}{ds}\right|^2 + \beta(s)\left|\frac{d^2(v(s))}{ds^2}\right|^2 \tag{5.10}$$

$$E_{snake} = \alpha(s)\left|\frac{d(v(s))}{ds}\right|^2 + \beta(s)\left|\frac{d^2(v(s))}{ds^2}\right| + \gamma(s)E_{image} \quad (5.11)$$

$$E_{image} = -\sum_i \left|G_x(x_i, y_i)\right|^2 + \left|G_y(x_i, y_i)\right|^2 \quad (5.12)$$

$$G_x = \frac{\partial}{\partial x} G_\sigma * I \quad (5.13)$$

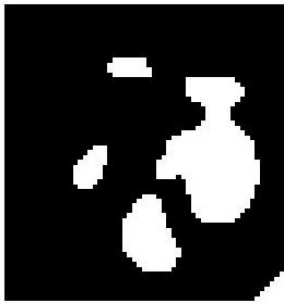
$$G_y = \frac{\partial}{\partial y} G_\sigma * I \quad (5.14)$$

Parameters $\alpha(s)$ and $\beta(s)$ are used to control the strength of the model's tension and rigidity respectively. The coefficient $\gamma(s)$ is introduced to make the units on the left side consistent with the right side. Thus, the minimization is solved by placing an initial contour on the image domain and allowing it to deform according to Equation 5.11. The initial contour is also named by mask and for the snake algorithm I used the result of the thresholding algorithm as a mask for perform this algorithm.

In the Figure 5.14 I observed the different masks used in the snake algorithm. The results of this algorithm can be seen in the Figure 5.15 and in the Table 5.7 the code used to process the snake algorithm on the images of the inner ear structures is shown.

(a)

(b)

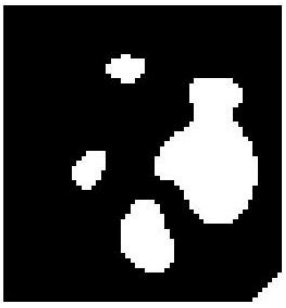


(c)

Figure 5.14: The different masks used to process the snake algorithms on the different images

(a)

(b)



(c)

Figure 5.15: Results of the snake algorithm

Table 5.7: Algorithm used to obtain the Figures 5.15

```
%Normalization
max_val = 1;
min_val = 0;
img_norm = normalize_img(roi_2, max_val, min_val);
```

```
%Create Mask and reverse the intensity of the image pixels
img_norm_1 = imadjust(img_norm,[0;1],[1;0]);
BW1 = im2bw(img_norm_1,0.5);
```

```
% Localized Region Based Active Contour Segmentation:
%
% seg = localized_seg(I,init_mask,max_its,rad,alpha,method)
%
% Inputs: I           2D image
%         init_mask   Initialization (1 = foreground, 0 = bg)
%         max_its     Number of iterations to run segmentation for
%         rad         (optional) Localization Radius (in pixels)
%                     smaller = more local, bigger = more global
%         alpha       (optional)  Weight of smoothing term
%                     higer = smoother
%         method      (optional) selects localized energy
%                     1 = Snake
%
%
% Outputs: seg        Final segmentation mask (1=fg, 0=bg)
%
%
% Description: This code implements the paper: "Localizing Region Based
% Active Contours" By Lankton and Tannenbaum.  In this work, typical
% region-based active contour energies are localized in order to handle
% images with non-homogeneous foregrounds and backgrounds.
%
%--------------------------------------------------------------------------

function seg =
localized_seg(I,init_mask,max_its,rad,alpha,method,FigRefreshRate,display)

  %-- default value for parameter alpha is .1
  if(~exist('alpha','var'))
    alpha = 0.05;
```

```
end
%-- default value for parameter method is 2
if(~exist('method','var'))
  method = 2;
end
%-- default behavior is to display intermediate outputs
if(~exist('display','var'))
  display = true;
end
 if(~exist('FigRefreshRate','var'))
  FigRefreshRate =20;
end
%-- Ensures image is 2D double matrix
%I = im2graydouble(I);
%-- Default localization radius is 1/10 of average length
[dimy dimx] = size(I);
if(~exist('rad','var'))
  rad = round((dimy+dimx)/(2*8));
  if(display>0)
    disp(['localiztion radius is: ' num2str(rad) ' pixels']);
  end
end

%-- Create a signed distance map (SDF) from mask
phi = mask2phi(init_mask);

%--main loop
for its = 1:max_its   % Note: no automatic convergence test

  %-- get the curve's narrow band
  idx = find(phi <= 1.2 & phi >= -1.2)';
  [y x] = ind2sub(size(phi),idx);

  %-- get windows for localized statistics
  xneg = x-rad; xpos = x+rad;      %get subscripts for local regions
  yneg = y-rad; ypos = y+rad;
  xneg(xneg<1)=1; yneg(yneg<1)=1;  %check bounds
  xpos(xpos>dimx)=dimx; ypos(ypos>dimy)=dimy;

  %-- re-initialize u,v,Ain,Aout
  u=zeros(size(idx)); v=zeros(size(idx));
  Ain=zeros(size(idx)); Aout=zeros(size(idx));

  %-- compute local stats
  for i = 1:numel(idx)  % for every point in the narrow band
    img = I(yneg(i):ypos(i),xneg(i):xpos(i)); %sub image
    P = phi(yneg(i):ypos(i),xneg(i):xpos(i)); %sub phi

    upts = find(P<=0);           %local interior
```

```
    Ain(i) = length(upts)+eps;
    u(i) = sum(img(upts))/Ain(i);

    vpts = find(P>0);            %local exterior
    Aout(i) = length(vpts)+eps;
    v(i) = sum(img(vpts))/Aout(i);
  end

  %-- get image-based forces
  switch method  %-choose which energy is localized
   case 1,               %-- Snake
    F = -(u-v).*(2.*I(idx)-u-v);
  end

  %-- get forces from curvature penalty
  curvature = get_curvature(phi,idx,x,y);

  %-- gradient descent to minimize energy
  dphidt = F./max(abs(F)) + alpha*curvature;

  %-- maintain the CFL condition
  dt = .45./(max(dphidt)+eps);

  %-- evolve the curve
  phi(idx) = phi(idx) + dt.*dphidt;

  %-- Keep SDF smooth
  phi = sussman(phi, .5);

  %-- intermediate output
  if((display>0)&&(mod(its,FigRefreshRate) == 0))
    showCurveAndPhi(I,phi,its);
  end
 end

 %-- final output
 if(display)
  showCurveAndPhi(I,phi,its);
 end

 %-- make mask from SDF
 seg = phi<=0; %-- Get mask from levelset
end
```

```
seg_z = localized_seg(img_norm_1, BW1, 300, 3, 0.05, 1);
```

Chan-Vese Model

The Chan-Vese is a geometric deformable model. The Chan-Vese model does not use the information of image gradient, but instead uses the intensity variations of foreground and background as the segmentation due. The energy functional of Chan-Vese model is defined as follows:

$$E = \mu.length(C) + v.Area(inside(C)) + \lambda_1 \int_{inside(C)} |I(x,y) - c_1|^2 dxdy + \lambda_2 \int_{outside(C)} |I(x,y) - c_2|^2 dxdy \tag{5.15}$$

where C represents the moving contour; I is the intensity function; $\mu, v \geq 0; \lambda_1, \lambda_2 > 0$ are the weights of each corresponding item; c_1 and c_2 are the mean of intensity values inside and outside the moving contour. The last two items are called the external energy, which is used to attract the contour to the correct position. Through the Euler-Lagrange equation of the functional, the moving equations derived as:

$$\frac{\partial \phi}{\partial t} = \delta_\varepsilon(\phi)\left[\mu.div\left(\frac{\nabla \phi}{|\nabla \phi|}\right) - v - \lambda_1 (I(x,y) - c_1)^2 + \lambda_2 (I(x,y) - c_2)^2 \right] \tag{5.16}$$

$$\phi(x,y,0) = \phi_0(x,y), \text{ in } \Omega \tag{5.17}$$

$$\frac{\delta_\varepsilon(\phi)}{|\nabla \phi|}\frac{\partial \phi}{\partial \vec{n}} = 0, \text{ on } \partial\Omega \tag{5.18}$$

where δ is the Dirac function; $\partial\phi / \partial\vec{n}$ denotes the normal derivate of ϕ at the boundary; Ω is the image region. The average intensities c_1 and c_2 can be calculated as:

$$c_1 = \frac{\int_{\Omega} I(x,y)H(\phi(x,y))dxdy}{\int_{\Omega} H(\phi(x,y))dxdy}, \text{ if } \int_{\Omega} H(\phi(x,y))dxdy > 0 \tag{5.19}$$

$$c_2 = \frac{\int_{\Omega} I(x,y)(1-H(\phi(x,y))dxdy}{\int_{\Omega} 1-H(\phi(x,y))dxdy}, \text{ if } \int_{\Omega} 1-H(\phi(x,y))dxdy > 0 \tag{5.20}$$

where H is the Heaviside function.

The Chan-Vese algorithm is exposed in the Table 5.8 and the results are illustrated in the Figure 5.16.

(a)

(b)

(c)

Figure 5.16: Results of the Chan-Vese Model in the region of interest

Table 5.8: The MATLAB code used to obtain segmented structures when is performed the chan-vese model

```
%Normalization
max_val = 1;
min_val = 0;
img_norm = normalize_img(roi_2, max_val, min_val);
```

```
%Create Mask and reverse the intensity of the image pixels
img_norm_1 = imadjust(img_norm,[0;1],[1;0]);
BW1 = im2bw(img_norm_1,0.5);
```

```
% seg = localized_seg(I,init_mask,max_its,rad,alpha,method)
%
% Inputs: I           2D image
%        init_mask   Initialization (1 = foreground, 0 = bg)
%        max_its     Number of iterations to run segmentation for
%        rad         (optional) Localization Radius (in pixels)
%                    smaller = more local, bigger = more global
%        alpha       (optional)  Weight of smoothing term
%                    higer = smoother
%        method         1 = Chan-Vese Energy
%
%
% Outputs: seg        Final segmentation mask (1=fg, 0=bg)
%
%
%
% Description: This code implements the paper: "Localizing Region Based
% Active Contours" By Lankton and Tannenbaum.  In this work, typical
% region-based active contour energies are localized in order to handle images with
non-homogeneous foregrounds and backgrounds.
%
%--------------------------------------------------------------------------

function seg =
localized_seg(I,init_mask,max_its,rad,alpha,method,FigRefreshRate,display)

 %-- default value for parameter alpha is .1
 if(~exist('alpha','var'))
   alpha = 0.05;
 end
 %-- default value for parameter method is 2
 if(~exist('method','var'))
   method = 2;
```

```
end
%-- default behavior is to display intermediate outputs
if(~exist('display','var'))
  display = true;
end
 if(~exist('FigRefreshRate','var'))
  FigRefreshRate =20;
end
%-- Ensures image is 2D double matrix
I = im2graydouble(I);
%-- Default localization radius is 1/10 of average length
[dimy dimx] = size(I);
if(~exist('rad','var'))
  rad = round((dimy+dimx)/(2*8));
  if(display>0)
    disp(['localiztion radius is: ' num2str(rad) ' pixels']);
  end
end

%-- Create a signed distance map (SDF) from mask
phi = mask2phi(init_mask);

%--main loop
for its = 1:max_its   % Note: no automatic convergence test

  %-- get the curve's narrow band
  idx = find(phi <= 1.2 & phi >= -1.2)';
  [y x] = ind2sub(size(phi),idx);

  %-- get windows for localized statistics
  xneg = x-rad; xpos = x+rad;      %get subscripts for local regions
  yneg = y-rad; ypos = y+rad;
  xneg(xneg<1)=1; yneg(yneg<1)=1;  %check bounds
  xpos(xpos>dimx)=dimx; ypos(ypos>dimy)=dimy;

  %-- re-initialize u,v,Ain,Aout
  u=zeros(size(idx)); v=zeros(size(idx));
  Ain=zeros(size(idx)); Aout=zeros(size(idx));

  %-- compute local stats
  for i = 1:numel(idx)  % for every point in the narrow band
    img = I(yneg(i):ypos(i),xneg(i):xpos(i)); %sub image
    P = phi(yneg(i):ypos(i),xneg(i):xpos(i)); %sub phi

    upts = find(P<=0);           %local interior
    Ain(i) = length(upts)+eps;
    u(i) = sum(img(upts))/Ain(i);

    vpts = find(P>0);            %local exterior
```

```
    Aout(i) = length(vpts)+eps;
    v(i) = sum(img(vpts))/Aout(i);
  end

  %-- get image-based forces
  switch method  %-choose which energy is localized
   case 1,              %-- CHAN VESE
    F = -(u-v).*(2.*I(idx)-u-v);
   end

  %-- get forces from curvature penalty
  curvature = get_curvature(phi,idx,x,y);

  %-- gradient descent to minimize energy
  dphidt = F./max(abs(F)) + alpha*curvature;

  %-- maintain the CFL condition
  dt = .45./(max(dphidt)+eps);

  %-- evolve the curve
  phi(idx) = phi(idx) + dt.*dphidt;

  %-- Keep SDF smooth
  phi = sussman(phi, .5);

  %-- intermediate output
  if((display>0)&&(mod(its,FigRefreshRate) == 0))
    showCurveAndPhi(I,phi,its);
  end
 end

 %-- final output
 if(display)
  showCurveAndPhi(I,phi,its);
 end

 %-- make mask from SDF
 seg = phi<=0; %-- Get mask from levelset
end
```

```
seg_y = localized_seg(img_norm_1, BW1, 300, 3, 0.05, 1);
```

Level set of the Li and Xu

Li & Xu (2005) define the following total energy functional

$$\varepsilon(\phi) = \mu \mathrm{P}(\phi) + \varepsilon_{g,\lambda,v}(\phi) \tag{5.21}$$

The external energy $\varepsilon_{g,\lambda,v}$ drives the zero level set toward the object boundaries, while the internal energy $\mu \mathrm{P}(\phi)$ penalizes the deviation of (ϕ) from a signed distance function during its evolution.

$$\varepsilon_{g,\lambda,v}(\phi) = \lambda L_g + v A_g(\phi) \tag{5.22}$$

In the equation (5.22) $\lambda > 0$ and v are constants, and the terms $L_g(\phi)$ and $A_g(\phi)$ are defined by:

$$L_g(\phi) = \int_{\Omega} g\delta(\phi)|\nabla\phi| dxdy \tag{5.23}$$

$$A_g(\phi) = \int_{\Omega} gH(-\phi) dxdy \tag{5.24}$$

Respectively, where δ is the univariate Dirac function, and H is the Heaviside function.

The g function is the edge indicator function defined by:

$$g = \frac{1}{1 + |\nabla G_\sigma * I|^2} \tag{5.25}$$

To understand the geometric meaning of the energy $L_g(\phi)$, Li & Xu (2005) suppose that the zero level set of ϕ can be represented by a differentiable parameterized curve $C(p)$, $p \in [0,1]$. It is well known that the energy functional $L_g(\phi)$ in equation (5.23)

computes the length of the zero level set curve of ϕ in the conformal metric $ds = g(C(p))|C'(p)|dp$. The energy functional $A_g(\phi)$ in equation (5.24) is introduced to speed up curve evolution. Note that, when the function g is constant 1, the energy functional in equation (5.24) is the area of the region $\Omega_\phi^- = \{(x,y) \mid \phi(x,y) < 0\}$. The energy functional $A_g(\phi)$ can be viewed as the weighted area of Ω_ϕ^-. The coefficient v of $A_g(\phi)$ can be positive or negative, depending on the relative position of the initial contour to the object of interest. For example, if the initial contours are placed outside the object, the coefficient v in the weighted area term should take positive value, so that the contours can shrink faster. If the initial contours are placed inside the object the coefficient v should take negative value to speed up the expansion of the contours.

By calculus of variations, the first derivate of the functional energy in equation (5.22) can be written as

$$\frac{\partial \varepsilon}{\partial \phi} = -\mu\left[\Delta\phi - div\left(\frac{\nabla\phi}{|\nabla\phi|}\right)\right] - \lambda\delta(\phi)div\left(g\frac{\nabla\phi}{|\nabla\phi|}\right) - vg\delta(\phi) \tag{5.26}$$

where Δ is the Laplacian operator. Therefore, the function ϕ that minimizes this functional satisfies the Euler-Lagrange equation $\frac{\partial \varepsilon}{\partial \phi} = 0$. The steepest descent process for minimization of the functional ε is the following gradient flow:

$$\frac{\partial \phi}{\partial t} = \mu\left[\Delta\phi - div\left(\frac{\nabla\phi}{|\nabla\phi|}\right)\right] + \lambda\delta(\phi)div\left(g\frac{\nabla\phi}{|\nabla\phi|}\right) + vg\delta(\phi) \tag{5.27}$$

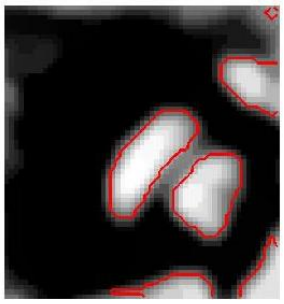
This gradient flow is the evolution equation of the level set function in the proposed method.

The second and the third term in the right hand side of equation (5.27) correspond to the gradient flows of the energy functional $\lambda L_g(\phi)$ and $vA_g(\phi)$, respectively, and are responsible of driving the zero level set curve towards the object boundaries. To explain the effect of the first term, which is associated to the internal energy $\mu P(\phi)$, Li & Xu (2005) notice that the gradient flow

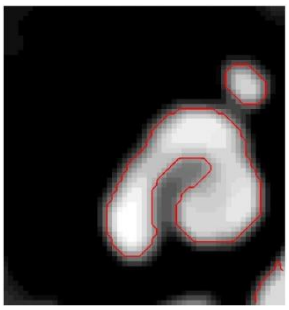
$$\Delta\phi - div\left(\frac{\nabla\phi}{|\nabla\phi|}\right) = div\left[\left(1 - \frac{1}{|\nabla\phi|}\nabla\phi\right)\right] \tag{5.28}$$

has the factor $\left(1 - \frac{1}{|\nabla\phi|}\right)$ as diffusion rate. If $|\nabla\phi| > 1$, the diffusion rate is positive and the effect of this term is the usual diffusion, making ϕ more even and therefore reduce the gradient $|\nabla\phi|$. If $|\nabla\phi| < 1$, the term has effect of reverse diffusion and therefore increase the gradient.

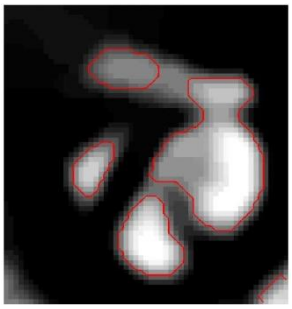
In the Figure 5.17 the results of the Li & Xu (2005) algorithm are expressed and the code is illustrated in the Table 5.9.



(a)



(b)



(c)

Figure 5.17: Results of the algorithm Li & Xu (2005)

Table 5.9: MATLAB code used to obtain the segmentation results

```
%Normalization
max_val = 1;
min_val = 0;
img_norm = normalize_img(roi_2, max_val, min_val);
img_norm_1 = imadjust(img_norm,[0;1],[1;0]);
mask = normalize_img(seg_y, max_val, min_val);
```

```
function
LevelSetEvolutionWithoutReinitialization(Img,mask,epsilon,mu,lambda,alf,c0,N,PlotR
ate,delt)
if(~exist('PlotRate','var'))
   PlotRate = 20;
end
[Ix,Iy]=gradient(Img);
f=sqrt(Ix.^2+Iy.^2);
g=1./(1+f);  % edge indicator function.
timestep=0.2/mu;
[nrow, ncol]=size(Img);
initialLSF= c0*2*(0.5-mask); % initial level set function: -c0 inside R, c0 outside R;
u=initialLSF;
imagesc(Img, [0, 1]);colormap(gray);hold on; axis off;axis equal;
contour(u,[0 0],'r','LineWidth',2);
title('Initial contour');

[vx,vy]=gradient(g);
lambda1 = 0.0005;
lambda2 = 0.0005;

% start level set evolution
for n=1:N
   u=NeumannBoundCond(u);
   [ux,uy]=gradient(u);
   normDu=sqrt(ux.^2 + uy.^2 + 1e-10);
   Nx=ux./normDu;
   Ny=uy./normDu;
   diracU=Dirac(u,epsilon);
   K=curvature_central(Nx,Ny);
   weightedLengthTerm=lambda*diracU.*(vx.*Nx + vy.*Ny + g.*K);
   penalizingTerm=mu*(4*del2(u)-K);
   weightedAreaTerm=alf.*diracU.*g;
   u=u+delt*(weightedLengthTerm + weightedAreaTerm + penalizingTerm);  % update
the level set function
end
```

```
if mod(n,PlotRate)==0
        pause(0.001);
        imagesc(Img, [0, 1]);colormap(gray);hold on;axis off;axis equal;
        contour(u,[0 0],'r','LineWidth',2);
        iterNum=['Level Set Evolution Without Re-initialization: A New Variational
Formulation ',num2str(n),' iterations'];
        title(iterNum);
        hold off;
end
imagesc(Img, [0, 1]);colormap(gray);hold on;
contour(u,[0 0],'r','LineWidth',2);
axis off;axis equal;
iterNum=['Level Set Evolution Without Re-initialization: A New Variational
Formulation ',num2str(n),' iterations'];
title(iterNum);
```

```
% the following functions are called by the main function EVOLUTION
function f = Dirac(x, sigma)
f=(1/2/sigma)*(1+cos(pi*x/sigma));
b = (x<=sigma) & (x>=-sigma);
f = f.*b;
end
```

```
function K = curvature_central(nx,ny)
[nxx,junk]=gradient(nx);
[junk,nyy]=gradient(ny);
K=nxx+nyy;
end
```

```
function g = NeumannBoundCond(f)
% Make a function satisfy Neumann boundary condition
[nrow,ncol] = size(f);
g = f;
g([1 nrow],[1 ncol]) = g([3 nrow-2],[3 ncol-2]);
g([1 nrow],2:end-1) = g([3 nrow-2],2:end-1);
g(2:end-1,[1 ncol]) = g(2:end-1,[3 ncol-2]);
end
```

```
%Li and Xu - Level set
delt= 0.001;
Img = img_norm_1;
epsilon = 1.5;
mu = 0.04;
alf = 1.5;
c0 = 4;
```

```
N = 1000;
PlotRate = 20;
lambda = 2;
gama = 0.6;

LevelSetEvolutionWithoutReinitialization(Img,mask,epsilon,mu,lambda,alf,c0,N,PlotR
ate,delt);
```

5.5 Algorithm Selection

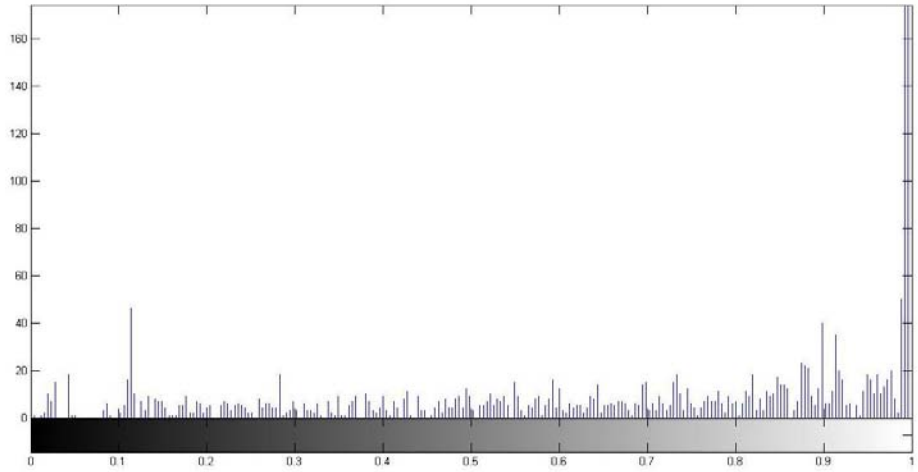
To select the best algorithm for realizing the segmentation of the inner ear structures will be analyzed the advantages, disadvantages and the obtained results of each one will be analyzed.

5.5.1 Otsu Method

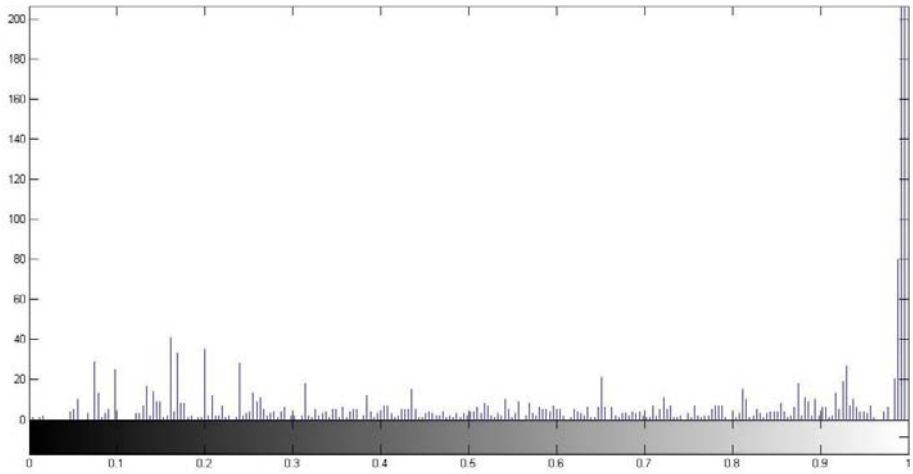
This method, not only finds the best threshold value in the statistical sense, but also it is the most stable method in the image thresholding segmentation at present. In addition, this method is good for thresholding a histogram with bimodal or multimodal distribution, in other words, it's optimal for thresholding large objects from the background. But it has difficulties processing images with unimodal or close to unimodal distributions. By histogram analysis of the regions of interest it is possible to see that the histograms have a distribution inappropriate for use the Otsu method, Figure 5.18. In the histograms a wrinkled modal characteristic are not observed.

Only the Figure 5.18.b observes modal properties between 0 - 0.4 and 0.8 - 1 intensity values.

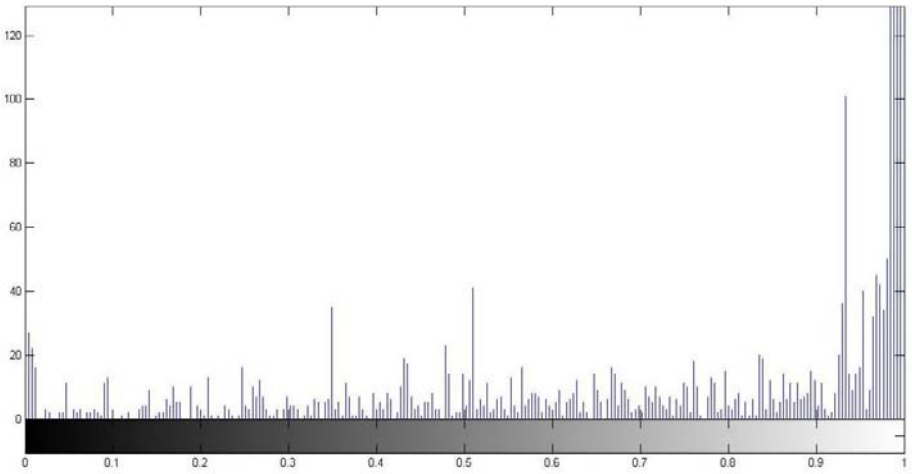
The main difficulties associated with thresholding are noise process, in the case of this be no-stationary, correlated and non-gaussian. Other factors complicating thresholding operation are ambient illumination, variance of gray levels within the object and the background, inadequate contrast, object shape and size non-commensurate with scene. Improper thresholding causes blotches, streaks and erasures on the image confounding segmentation and recognition tasks. In some cases, the merges, fractures and other deformations in the character shapes as a consequence of incorrect thresholding are known to be the main reasons of optical character recognition performance deteoration.



(a)



(b)



(c)

Figure 5.18: Histograms of the different region of interest figures

In the specific case, in the Figure 5.9.c a part of semicircular canal is not present, so this method, for this specific image, makes to eliminate some important structures of the image. In the other two results, Figure 5.9.a-b, it's possible to observe the main structures of the inner ear and from these results is more easy to understand why this method is one of the most used in these type of images, however thresholding algorithm depends on a multitude of a factors, such as, the gray level distribution, local shading effects and the presence of denser.

5.5.2 Region Growing

The results of region growing depend strongly on the selection of the homogeneity criterion. If the homogeneity criterion is not properly chosen, the region leak out into adjoining areas, or merge with regions that do not belong to the object of interest. Therefore, region growing is seldom used alone. The primary disadvantage of the region growing is that it requires manual interaction to obtain the seed point. Thus, for each region that needs to be extracted, a seed must be planted. In Figure 5.11.c one structure of the semicircular canal is not present, because in the image a seed was not planted in the local of the structure.

Another problem of the region growing is that different starting points may not grow into identical regions. Otherwise, the advantage of region growing is that it is capable of correctly segmenting regions that have the same properties and are spatially separated. Another advantage is that it generates connected regions; however, region growing can also be sensitive to noise, make extracted regions to have holes or even become disconnected. Conversely, partial volume effects can cause separate regions to become connected. Considering that the CT images are really affected by partial volume effects and given that this method is very dependent of the manual interaction and of the localization of seeds this method cannot be considered very effective. In addition, the gray level distribution of the CT images is very large, ergo, it is more difficult to obtain a really effective result.

5.5.3 Canny Operator

The canny operator is characterized by a good localization and a single response to an edge, but the boundaries obtained by the Canny operator in the Figure 5.12.a-c are discontinuous. The Canny operator is known to present edges discontinuity due to the

noises, to the partial volume effects and to the very different orientation of the boundary pixels. The spatial relationships of the edge points are not reflected; as such, most of the detected boundaries are incomplete or wrongly connected.

The noise effects and the discontinuity of the boundary edges are two aspects that indicate a dependent algorithm of the pre and post processing.

5.5.4 Watershed

The watershed algorithm presents advantages and disadvantages. An advantage of the watershed method is that it can segment multiple objects in a simple thresholding setting. Otherwise, the disadvantage of the watershed method is that the different types of images need different thresholds and if the thresholds are not set correctly, then the objects are under and over segmented. The under and over segmentation is presented in the Figure 5.13.a-c. Furthermore, slight changes in the threshold can significantly alter the segmentation results.

5.5.5 Snake

The snake method is the first deformable model applied to the medical image segmentation and the development of parametric deformable models has a tight relationship with the snake method. The original snake method used the tension and rigidity of the contour, the internal energy, the gradient magnitude as the external energy. Furthermore, the snake method is sensitive to the initial conditions. The moving contour may stop at places with local functional minimum or places where the gradient magnitude is too small so that the external forces tend to be zero. In the non-interactive applications, the snakes must be initialized close to the structure of interest to guarantee good performance. Consequently, in order to get a correct segmentation the initial contour must have the same topology as the desired object and must be placed near the object boundary and Figure 5.14.a-c represents the initial contour and it has the same topology as the desired object. Thus, there are two key difficulties with parametric active contour algorithms. First, the initial contour must, in general, be close to the true boundary or else it will likely converge to the wrong result. The basic idea to address this problem is to increase the capture range of the external force fields and to guide the contour toward the desired boundary. The second problem is that active contours have difficulties processing into boundary concavities.

In the Figure 5.15.a-c the segmentation results of the snake algorithm is observed. In this case, the snake algorithm has poor convergence to boundaries with larger curvatures. Additionally, the algorithm performance has a high dependence on the initial contour.

5.5.6 Chan-Vese Model

The main idea of the level set method is to implicitly embed the moving contour into a higher dimensional level set function and view the contour as its zero level set. Then instead of tracking the discrete contour points, one can track the zero level set of the level set function. The advantage of doing so is that the topological changes can be easily handled and the geometric properties of the contour can be implicitly calculated. Therefore, the complexity of geometric deformable models, speed functions should be defined properly to drive the contour to the right position. Usually, the deformable models implemented by means of the level set method suffer from a slower speed of convergence than parametric deformable models due to their computational complexity. However, they can automatically handle topology changes and allow for multiple simultaneous boundary estimations. Specifically, algorithms based on geometric deformable models aim to eliminate noise influence, prevent leakage, enhance accuracy and efficiency, and make the algorithms more automatic and less dependent on the initial conditions.

The Figure 5.16.a-c is showing the results of the Chan-Vese model and when compared with snake algorithm is observed more efficiency in the convergence of the contour and the contour topology is really similar to the real boundary of interest object.

5.5.7 Level set of Li & Xu (2005)

The geometric deformable models are promising for the segmentation of the medical images because these models can easily incorporate statistical information and other techniques, while using curve evolution to find the optimal boundaries can provide a contour with regular geometric properties, and the Li & Xu (2005) algorithm has this properties / qualities.

In the Figure 5.17.a-c the results of the Li & Xu (2005) algorithm is illustrated and when it is compared the Chan-Vese model with the Li & Xu (2005) algorithm it is possible conclude that the last is less influenced by noise and the regulating effects of the internal forces make the boundary shape more reasonable.

5.5.8 Selection

When compared the thresholding algorithms with deformable models in the generalized characteristics it is possible to know that the thresholding algorithms present the following advantages and disadvantages:

Advantages:

- Simple;
- Effective;
- Low computational cost;
- Low processing time.

Disadvantages:

- Sensitive to partial volume effects;
- Don’t present good results when used in multi-channel images;
- The elimination noise presents a difficult control;
- In the case of region-growing is necessary to select a seed point;
- Appearance of holes and discontinuities;
- Slight changes in the value of threshold may lead to different results.

The deformable models method also presents a list of advantages and disadvantages:

Advantages:

- Simple;
- Exact;
- Accessible;
- Don’t depend of subjective parameters;
- Low sensitive to artifacts;
- Easy handling of shape variation;
- Mainly used for segmentation of structures of high complexity;
- Directly produces closed parametric curves and surfaces.

Disadvantages:

- In the initial stage is necessary the interaction of operator for placing the initial model and to choice the input parameters.

Through analyzing the lists of advantages and disadvantages that each type of method presents, it is easy to conclude that deformable models method presents more advantages than thresholding algorithms. Additionally, deformable models are more indicated for CT images than thresholding algorithms because of the high distribution in the gray level values of the pixels.

When the results of the thresholding-based algorithms are observed, it is possible see that the boundaries of the Canny edge detector results are discontinuous and the spatial relationship of the edge points are not reflected; such as, most of the detected boundaries are incomplete or wrongly connected. The watershed algorithm gives a complete segmentation of the image, but, over segmentation can be seen because there are a lot of pixels with local maximum of gradient magnitude. Finally, the area of the segmented objects by the Otsu method is excessively large.

Analyzing results allow concluding that in the figures, which represent results of deformable models, it is possible to see more structures of the inner ear than in the results of the thresholding algorithms. Furthermore, if you compare the three types of deformable models, the last two appear more effective in the representation of the real topology and shape of the inner ear structures. The last two are the Chan-Vese model and Li & Xu algorithm. The both of them are more convergent and less influenced by the noise / artifacts and initial contours.

Thus, I selected the geometric deformable models as the best algorithm for this type of images and to segment small structures of the inner ear.

5.6 Summary

Through this last study, I concluded that the type of images most commonly used are the CT images, and it is an added value because I was going to work with this material type. Another conclusion, which came from the same study, was the observation of the main used algorithms. The thresholding-based algorithms and deformable models are the main algorithms and they also are the most used.

The Function enhancement consists into three stages: first it was used the histogram stretch, second the Gaussian filtering was performed and finally, it was filtered by the Anisotropic Diffusion filter. To reduce the required computational time and also the

computational cost, the region of the inner ear, i.e. region of interest (ROI), was first selected from the filtered image by searching for the pixels with highest intensity values. The membranous labyrinth of the inner ear is fully enclosed by the resultant ROI. Then, from membranous labyrinth, it is possible to create a region that represents the boundary of the temporal bone. By this way the image size is reduced and the structures of interest are fully inside the ROI.

Once known the type of segmentation algorithm most used and with the enhanced ROI, I performed experimental tests and I used the Otsu method, region growing algorithm, canny operator and watershed to analyze the results for a variety of thresholding-based algorithms. I also made the same for deformable models, but for this specific case, I realized the snake, the Chan-Vese model and the Li & Xu (2005) level set algorithm.

In the finally stage, I was selected the best type of segmentation algorithm and through the analysis of the results and by study of advantages and disadvantages of each one was concluded that the deformable models seem to be promising for the segmentation of the inner ear because they can easily incorporate statistical information in order to improve their performance / efficiency / effectiveness.

Chapter VI - Final Conclusions and Future Work

- Final Conclusions;
- Future Perspectives.

6.1 Final Conclusions

In this Dissertation I present a study for to choose a suitable algorithm to segment the inner ear structures by analyzing medical images, in the specific case, Computerized Tomography images. The used images are more appropriate to visualize the structures of the middle ear and, because of this reason; I have only analyzed 3 slices that clearly demonstrate the presence of the inner ear structures.

In more detail, during this study, the following tasks were concluded:

- A review on the segmentation algorithms commonly used in medical images.
- A review in the current works concerning the segmentation of ear structures.
- The development of a very effective method of image preprocessing that driven the segmentation to the region in which are contained the structures of interest, and simultaneously decreases the computational cost and time processing. As such, the segmentation results are obtained almost immediately.
- The identification of the method that presents the best segmentations of the inner ear.

From the review studies, it was realized that:

- The common image segmentation algorithms are often divided into: thresholding-based, clustering-based and deformable models.

- The types of images most commonly used to study the inner ear are the CT images.

- The thresholding-based algorithms and deformable models are the approaches most used to segment the inner ear.

The preprocessing algorithm developed is constituted by three phases:

- Read the input image;
- Enhancement of the input image;
- Definition of the region of interest.

The image enhancement consists into three stages: firstly, it used histogram stretch, secondly the Gaussian filtering was performed and finally, it was filtered by the Anisotropic Diffusion filter. To reduce the required computational time and also the computational cost, the region of the inner ear, i.e. region of interest (ROI), is defined in the filtered image by searching for the pixels with highest intensity values. The membranous labyrinth of the inner ear is fully enclosed by the resultant ROI. Then, from the membranous labyrinth, it is possible to create a region that represents the boundary of the temporal bone. By this way, the image size is reduced and the structures of interest are fully contained into the ROI.

In the experimental tests done, the Otsu method, region growing algorithm, Canny operator and watershed were used to analyze the results obtained by a variety of thresholding-based algorithms, and also by deformable models, like the snake model, the Chan-Vese's model and the Li & Xu's level set method.

Then, the experimental results were evaluated making possible to realize that deformable models seem to be promising for the segmentation of the inner ear, because they can easily incorporate statistical information in order to improve their effectiveness.

Also, from the experimental results, it was possible to conclude that the deformable models were able to segment more structures of the inner ear than the thresholding-based algorithms. Furthermore, when compared the three types of deformable models used, the level set method was more effective in the segmentation of the real topology and shape of the inner ear structures. The Chan-Vese's model and the Li & Xu (2005) level set method were more convergent and less influenced by noise, artifacts and initial contours. The level set method used is a geometric deformable model that can automatically handle topology changes and allow simultaneously multiple boundary segmentations. Specifically, algorithms based on geometric deformable models aim to eliminate noise influence, prevent leakage, enhance accuracy and efficiency, and make the algorithms more automatic and less dependent on the initial contour. This was been confirmed by the experimental tests done.

6.2 Future Perspectives

Despite the comprehensive approach adopted and implemented throughout this Dissertation that produced experimental results quite satisfactory and promising, the work done can be enriched and improved, in particular, by addressing the following points:

- Enrichment and improvement of the study done by performing a statistical analysis of the results obtained using images manually segmented by experts as ground-truth.
- Use of images specially acquired for the inner ear, since the used images are specific to the middle ear and consequently the inner ear is not represented effectively.
- 3D visualization of the segmented contours and 3D reconstruction of the structures segmented by interpolating those contours.
- Building of biomechanical models for the main auditory organs from the 3D models built, that can be used to assist cochlear implant surgery.

References

Akhtar, P. and T. J. Ali (2008). Edge detection and linking using wavelet representation and image fusion. Image and Signal Processing, 2008. CISP '08. Congress Sanya, Hainan**:** 273-277.

Bagirov, A. M., J. Ugon, et al. (2011). "Fast modifiedglobal k-means algorithmforincrementalclusterconstruction." Pattern Recognition **44**: 866-876.

Bai, X., F. Zhou, et al. (2009). "Enhanced detectability of point target using adaptive morphological clutter elimination by importing the properties of the target region." Signal Processing **89**: 1973-1989.

Bankman, I. N. (2000). Handbook of Medical Imaging Processing and Analysis. San Diego, Academic Press.

Beevi, S. Z. (2010). "An Effective Approach for Segmentation of MRI Images: Combining Spatial Information with Fuzzy C-Means Clustering." European Journal of Scientific Research **41**: 437-451.

Betanzos, A. A., B. A. Varela, et al. (2000). "Analysis and evaluation of hard and fuzzy clustering segmentation techniques in burned patient images." Image and Vision Computing **18**: 1045-1054.

Bosworth, J. H. and S. T. Acton (2003). "Morphological scale-space in image processing." Digital Signal Processing **13**: 338-367.

Bradshaw, A. P., I. S. Curthoys, et al. (2010). "A Mathematical Model of Human Semicircular Canal Geometry: A New Basis for Interpreting Vestibular Physiology." Journal of the Association for Research in Otolaryngology **11**: 145-159.

Bruijne, M. d., B. v. Ginneken, et al. (2002). "Active shape model based segmentation of abdominal aortic aneutysms in CTA images." Proceedings of SPIE **4684**: 463-474.

Calhoun, P. S., B. S. Kuszyk, et al. (1999). "Fast Surface and Volume Rendering Based on Shear-

Warp Factorization for a Surgical Simulator." RadioGraphics **19**: 745-764.

Cannon, K. (2007). "Hearinginfo.org." from http://www.hearinginfo.org/diseasesdisorders/otosclerosis.htm.

Carr, N. (2010). "Middle Ear." Retrieved 22 - 11 - 2010, 2010.

Chen, W.-g. (2006). "Gradient vector flow using an implicit method." International Journal of Information Technology **12**: 14-23.

Chen, Y. B. and O. T. C. Chen (2009). "Image Segmentation Method Using Thresholds AUtomatically Determined from Picture Contents." EURASIP Journal on Image and Video Processing **2009**: 15.

Christensen, G. E., J. He, et al. (2003). "Automatic Measurement of the Labyrinth Using Image Registration and a Deformable Inner Ear Atlas." Academic Radiology Journal **10**: 988-999.

Comunello, E., A. v. Wangenheim, et al. (2009). "A computational method for the semi-automated quantitative analysis of tympanic membrane perforations and tympanosclerosis." Computers in Biology and Medicine **39**: 889-895.

Davis, A. (2010). "Hear - More and more hearing impaired people." Retrieved 26 - 11 - 2010, 2010, from http://www.hear-it.org/page.dsp?area=134.

Decraemer, W. F., J. J. J. Dirckx, et al. (2003). "Three-Dimensional Modelling of the Middle-Ear Ossicular Chain Using a Commercial High-Resolution X-Ray CT Scanner." Journal of the Association for Research in Otolaryngology **4**: 250-263.

Dempster, A. P., N. M. Laird, et al. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm." Journal of the Royal Statistical Society **39**: 1-38.

Ding, L. and A. Goshtasby (2001). "On the Canny edge detector." Pattern Recognition **34**: 721-725.

Duchesne, S., J. C. Pruessner, et al. (2002). "Appearance-Based Segmentation of Medial Temporal Lobe Structures

" NeuroImage **17**: 515-531.

Fang, M., G. X. Yue, et al. (2009). The study on application of Ostu Method in Canny Operator. Proceedings of the 2009 International Symposium on information Processing, Huangshan, P. R. China.

Figueiredo, M. A. T. (2007). Semi-Supervised Clustering: Application to Image Segmentation. Advances in Data Analysis R. Decker and H.-J. Lenz. Berlin, Springer Berlin Heidelberg**:** 39-50.

Glynn, E. F. (2007). "Mixture of Gaussian Distributions." Retrieved 12-01-2011, 2011, from http://research.stowers-institute.org/efg/R/Statistics/MixturesOfDistributions/index.htm.

Gray, H. (1985). Anatomy of the Human Body: Gray's Anatomy. Baltimore, Williams & Wilkins

Hain, T. C. (2010). "Hearing Loss." Retrieved 25 - 11 - 2010, 2010, from http://www.dizziness-and-balance.com/disorders/hearing/hearing.html.

Henrique, L. L. (2002). Acústica Musical, Fundação Calouste Glubenkian.

Hussong, A., T. S. Rau, et al. (2009). "An automated insertion tool for cochlear implants: another step towards atraumetic cochlear implant surgery." Int Journal CARS **5**: 163-171.

Irwin, J. (2006). Basic Anatomy and Physiology of the Ear. Infection and Hearing Impairement. V. E. N. a. P. J. Vallely, John Wiley & Sons, Ltd.

Jun, B.-C., S.-W. Song, et al. (2005). "Three-dimensional reconstruction based on images from spiral high-resolution computed tomography of the temporal bone: anatomy and clinical application." The Journal of Laryngology & Otology **119**: 693-698.

Kaneshiro, N. K. (2010). "MedlinePlus Trusted Health Information for You." Retrieved 25 - 11 - 2010, 2010, from http://www.nlm.nih.gov/medlineplus/ency/article/007322.htm.

Kanopoulos, N., N. Vasanthavada, et al. (1988). "Design of an Image Edge Detection Filter Using the Sobel Operator

" IEEE Journal of Solid-State Circuits **23**: 358-367.

Kim, K. H., M. J. Kwon, et al. (2002). "Fast Surface and Volume Rendering Based on Shear-

Warp Factorization for a Surgical Simulator." Computer Aided Surgery **7**: 268-278.

Kobbelt, L. P., J. Vorsatz, et al. (1999). "A Shrink Wrapping Approach to Remeshing Polygonal Surfaces." EUROGRAPHICS **18**.

Koo, B. K., Y. K. Choi, et al. (2007). "Shrink-Wrapped Boundary Face Algorithm for Mesh Reconstruction from Unorganized Points." ETRI Journal **27**.

Lane, J. I., R. J. Witte, et al. (2005). "Imaging Microscopy of the Middle and Inner Ear Part II: MR Microscopy." Clinical Anatomy Wiley-Liss **18**: 409-415.

Lee, D. H., S. Chan, et al. (2010). "Reconstruction and exploration of virtual middle-ear models derived from micro-CT datasets." Hearing Research **263**: 198-203.

Li, Y. and Y. Shen (2010). "Fuzzy c-means clustering based on spatial neighborhood information for image segmentation." Journal of Systems Engineering and Electronics **21**: 323-328.

Liu, B., X. L. Gao, et al. (2007). "A detailed 3D model of the guinea pig cochlea." Brain Struct Funct **212**: 212-230.

Ma, Z., J. M. R. S. Tavares, et al. (2010). "A review of algorithms for medical image segmentation and their applications to the female pelvic cavity." Computer Methods in Biomechanics and Biomedical Engineering **13**: 235-246.

Manousakas, I. N., P. E. Undrill, et al. (1998). "Split-and-Merge Segmentation of Magnetic Resonance Medical Images: Performance Evaluation and Extension to Three Dimensions." Computers and Biomedical Research **31**: 393-412.

Martin, F. N. and J. G. Clark (2011). Introduction to Audiology Allyn & Bacon.

Matta, F. and J.-L. Dugelay (2009). "Person recognition using facial video information: state of the art." Journal of Visual Languages and Computing **20**: 180-187.

McAullife, M. (2010). "Edge Detection: Zero X Laplacian." Medical Image Processing Analysis, & Visualization Retrieved 13-01-2011, 2011, from http://mipav.cit.nih.gov/documentation/HTML%20Algorithms/Edge_Detection_Zero_X.html.

McInerney, T. and D. Terzopoulos (1996). "Deformable Models in Medical Image Analysis: A Survey." Medical Image Analysis **2**: 91-108.

Melhem, E. R., H. Shakir, et al. (1998). "Inner Ear Volumetric Measurements Using High-Resolution 3D T2-Weighted Fast Spin-Echo MR Imaging: Initial Experience in Healthy Subjects." American Journal Of Neuroradiol **19**: 1819-1822.

Mille, J. (2009). "Narrow band region-based active contours and surfaces for 2D and 3D segmentation " Computer Vision and Image Understanding **113**: 946-965.

Mlsna, P. A. and J. J. Rodríguez (2009). Gradient and Laplacian Edge Detection. The Essential Guide to Image Processing. Elsevier. San Diego**:** 495-524.

Moller, A. R. (2006). Hearing: Anatomy, Physiology, and Disorders of the Auditory System, Second Edition. San Diego, Academic Press.

Newman, T. S. and H. Yi (2006). "A survey of the marching cubes algorithm." Computers & Graphics **30**: 854-879.

Noble, J. H., F. M. Warren, et al. (2008). "Automatic segmentation of the facial nerve and chorda tympani in CT images using spatially dependent feature values." Medical Physics **35**: 5375-5384.

Paiva, A. R. C. and T. Tasdizen (2010). Fast Semi-Supervised Image Segmentation by Novelty Selection. Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference Salt Lake City, IEEE**:** 1054 - 1057

Papari, G. and N. Petkov (2011). "Edge and line oriented contour detection: State of the art." Image and Vision Computing **29**: 79-103.

Pham, D. L., C. Xu, et al. (2000). "Current Methods in Medical Image Segmentation." Annu. Rev. Biomed. Eng. **2**: 315-337.

Poznyakovskiy, A. A., T. Zahnert, et al. (2008). "The creation of geometric three-dimensional models of the inner ear based on micro computer tomography data." Hearing Research **243**: 95-104.

Rau, T. S., A. Hussong, et al. (2010). "Automated insertion of performed cochlear implant electrodes: evaluation of curling behaviour and insertion forces on an artificial cochlear model." Int Journal CARS **5**: 173-181.

Rodt, T., P. Ratiu, et al. (2002). "3D visualisation of the middle ear and adjacent structures using reconstructed multi-slice CT datasets, correlating 3D images and virtual endoscopy images." Neuroradiology **44**: 783-790.

Seeley, Stephens, et al. (2004). The Special Senses, The MacGraw-Hill Companies.

Seemann, M. D., O. Seemann, et al. (1999). "Evaluation of the middle and inner ear structures: comparison of hybrid rendering, virtual endoscopy and axial 2D source images." European Radiology **9**: 1851-1858.

Shi, L., D. Wang, et al. (2010). "Automatic MRI segmentation and morphoanatomy analysis of the vestibular system in adolescent idiopathic scoliosis." NeuroImage: 9.

Sim, J. H. and S. Puria (2008). "Soft Tissue Morphometry of the Malleus-Incus Complex from Micro-CT Imaging." Journal of the Association for Research in Otolaryngology **9**: 5-21.

Sutton, M. A., J. C. Bezdek, et al. (2000). Image Segmentation by Fuzzy Clustering: Methods and Issues. Handbook of Medical Imaging Processing and Analysis. San Diego, Academic Press Series in Biomedical Engineering.

Todd, C., M. Tarabichi, et al. (2009). A computer-based, Interactive Tool for Semi-Automatic Extraction, Visualization and Pre-operative Assessment of the Inner Ear. Joint Conference on Medical Informatics. Taiwan, Taipei.

Tsai, A. Y. A. and A. S. Willsky (2001). "Curve Evolution Implementation of the Mumford-Shah Functional for Image Segmentation, Denoising, Interpolation, and Magnification." IEEE transactions on Image Processing **10**: 1169-1186.

Tuck-Lee, J. P., P. M. Pinsky, et al. (2008). "Finite element modeling of acoustical-mechanical coupling in the cat middle ear." Journal Acoustical Society of America **124**: 348-362.

UMMC. (2011). "Hearing and Balance Center." from http://www.umm.edu/otolaryngology/ear_infections.htm.

Vignoles, G. L., M. Donias, et al. (2010). "Simplified marching cubes: An efficient discretization scheme for simulations of deposition/ablation in complex media." Computational Materials Science.

Wang, Y., Y. Zhu, et al. (2007). Medical image segmentation based on deformable models and its applications. Deformable Models Theory and Biomaterial Applications, Springer New York**:** 209-260.

Wismuller, A., F. Vietze, et al. (2000). Segmentation with Neural Networks. Handbook of Medical Imaging Processing and Analysis. San Diego, Academic Press Series in Biomedical Engineering.

Xianfen, D., C. Siping, et al. (2005). 3D Semi-automatic Segmentation of the Cochlea and Inner Ear. Engineering in Medicine and Biology 27th Annual Conference. IEEE. Shangai, China.

Xiao-Juan, C. and L. Dan (2010). Medical image segmentation based on threshold SVM. ICBECS2010 Conference Schedule. IEEE. China.

Xiao-min, B., P. Xiao, et al. (2009). Textile Image Segmentation Based on Semi-supervised Clustering and Bayes

Decision. International Conference on Artificial Intelligence and Computational Intelligence, IEEE**:** 559-562.

Xie, J. and H. T. Tsui (2004). "Image segmentation based on maximum-likelihood estimation and optimum entropy-distribution (MLE-OED)." Pattern Recognition Letters **25**: 1133-1141.

Xie, X., M. Mirmehdi, et al. (2005). "Detecting Abnormalities in Tympanic Membrane Images." Medical Image Understanding and Analysis: 19–22.

Xu, C. and J. L. Prince (1998). "Snakes, Shapes, and Gradient Vector Flow." IEEE Transactions on Image Processing **7**: 359-369.

Yoo, S. K., G. Wang, et al. (2001). "Semiautomatic Segmentation of the Cochlea Using Real-Time Volume Rendering and Regional Adaptative Snake Modeling." Journal of Digital Imaging **14**: 173-181.

Zhao, Z., Z. Su, et al. (2006). "SVM for density estimation and application to medical image segmentation." Journal of Zhejiang University Science B **5**: 365-372.

Zhu, Y. (2010). An Efficient Supervised Clustering Algorithm Based on Neural Networks. 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE). IEEE. **4:** 265-268.